



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K 14/435, C12N 1/00, 1/10, 5/10, 15/12, 15/63		A1	(11) International Publication Number: WO 00/34326 (43) International Publication Date: 15 June 2000 (15.06.00)
(21) International Application Number: PCT/US99/29473 (22) International Filing Date: 10 December 1999 (10.12.99)		(74) Agent: ADLER, Benjamin, A.; McGregor & Adler, 8011 Candle Lane, Houston, TX 77071 (US).	
(30) Priority Data: 09/210,330 11 December 1998 (11.12.98) US 09/418,529 9 December 1999 (09.12.99) US		(81) Designated States: JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(71) Applicant (for all designated States except US): CLONTECH LABORATORIES, INC. [US/US]; 1020 East Meadow Drive, Palo Alto, CA 94303 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): LUKYANOY, Sergey Anatolievich [RU/RU]; ul. Golubinskaya 13/1-161, Moscow (RU). FRADKOV, Arcady Fedorovich [RU/RU]; ul. Dnepropetrovskaya, 35/2-14, Moscow, 113570 (RU). LABAS, Yulii Aleksandrovich [RU/RU]; ul. Generala Tyuleneva, 35-416, Moscow, 117465 (RU). MATZ, Mikhail Vladimirovich [RU/RU]; ul. Teplii stan 7/2-28, Moscow, 117465 (RU). GREEN, Gisele [US/US]; 4106 Thain Way, Sunnyvale, CA 94087 (US). CHEN, Ying [CN/US]; 680 Garland Avenue, #6, Sunnyvale, CA 94086 (US). DING, Li [CN/US]; 1352 Norman Drive, Sunnyvale, CA 94087 (US).		Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(54) Title: FLUORESCENT PROTEINS FROM NON-BIOLUMINESCENT SPECIES OF CLASS ANTHOZOA, GENES ENCODING SUCH PROTEINS AND USES THEREOF			
(57) Abstract The present invention is directed to novel fluorescent proteins from non-bioluminescent organisms from the Class Anthozoa. Also disclosed are cDNAs encoding the fluorescent proteins.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**FLUORESCENT PROTEINS FROM NON-BIOLUMINESCENT SPECIES
OF CLASS ANTHOZOA, GENES ENCODING SUCH PROTEINS AND
5 USES THEREOF**

10

BACKGROUND OF THE INVENTION

Cross-reference to Related Application

This is a divisional application of U.S.S.N. 09/210,330 filed
15 on December 11, 1998.

Field of the Invention

This invention relates to the field of molecular biology. More specifically, this invention relates to novel fluorescent proteins, 20 cDNAs encoding the proteins and uses thereof.

Description of the Related Art

Fluorescence labeling is a particularly useful tool for marking a protein, cell, or organism of interest. Traditionally, a 25 protein of interest is purified, then covalently conjugated to a fluorophore derivative. For *in vivo* studies, the protein-dye complex is then inserted into cells of interest using micropipetting or a method of reversible permeabilization. The dye attachment and insertion steps, however, make the process laborious and difficult to control. An

alternative method of labeling proteins of interest is to concatenate or fuse the gene expressing the protein of interest to a gene expressing a marker, then express the fusion product. Typical markers for this method of protein labeling include β -galactosidase, firefly luciferase 5 and bacterial luciferase. These markers, however, require exogenous substrates or cofactors and are therefore of limited use for *in vivo* studies.

A marker that does not require an exogenous cofactor or substrate is the green fluorescent protein (GFP) of the jellyfish 10 *Aequorea victoria*, a protein with an excitation maximum at 395 nm, a second excitation peak at 475 nm and an emission maximum at 510 nm. GFP is a 238-amino acid protein, with amino acids 65-67 involved in the formation of the chromophore.

Uses of GFP for the study of gene expression and protein 15 localization are discussed in detail by Chalfie et al. in *Science* 263 (1994), 802-805, and Heim et al. in *Proc. Nat. Acad. Sci.* 91 (1994), 12501-12504. Additionally, Rizzuto et al. in *Curr. Biology* 5 (1995), 635-642, discuss the use of wild-type GFP as a tool for visualizing 20 subcellular organelles in cells, while Kaether and Gerdes in *Febs Letters* 369 (1995), 267-271, report the visualization of protein transport along the secretory pathway using wild-type GFP. The expression of GFP in plant cells is discussed by Hu and Cheng in *Febs Letters* 369 (1995), 331-334, while GFP expression in *Drosophila* embryos is described by Davis et al. in *Dev. Biology* 170 (1995), 726-729.

25 Crystallographic structures of wild-type GFP and the mutant GFP S65T reveal that the GFP tertiary structure resembles a barrel (Ormo et al., (1996) *Science* 273: 1392-1395; Yang, et al., (1996) *Nature Biotech* 14: 1246-1251). The barrel consists of beta sheets in a compact antiparallel structure, where, in the center, an alpha helix

containing the chromophore is shielded by the barrel. The compact structure makes GFP very stable under diverse and/or harsh conditions such as protease treatment, making GFP an extremely useful reporter in general. However, the stability of GFP makes it sub-optimal for 5 determining short-term or repetitive events.

A great deal of research is being performed to improve the properties of GFP and to produce GFP reagents useful and optimized for a variety of research purposes. New versions of GFP have been developed, such as a "humanized" GFP DNA, the protein product of 10 which has increased synthesis in mammalian cells (Haas, et al., *Current Biology* 6 (1996), 315-324; Yang, et al., *Nucleic Acids Research* 24 (1996), 4592-4593). One such humanized protein is "enhanced green fluorescent protein" (EGFP). Other mutations to GFP have resulted in blue-, cyan- and yellow-green light emitting versions. Despite the great 15 utility of GFP, however, other fluorescent proteins with properties similar to or different from GFP would be useful in the art. Novel fluorescent proteins result in possible new colors, or produce pH-dependent fluorescence. Other benefits of novel fluorescent proteins include fluorescence resonance energy transfer (FRET) possibilities 20 based on new spectra and better suitability for larger excitation.

The prior art is deficient in novel fluorescent proteins wherein the DNA coding sequences are known. The present invention fulfills this long-standing need in the art.

25

SUMMARY OF THE INVENTION

The present invention is directed to DNA sequences encoding fluorescent proteins selected from the group consisting of:

(a) an isolated DNA from an organism from the Class Anthozoa which encodes a fluorescent protein; (b) an isolated DNA which hybridizes to the isolated DNA of (a) and which encodes a fluorescent protein; and (c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to the degeneracy of the genetic code and that encodes a fluorescent protein. Preferably, the DNA is isolated from a non-bioluminescent organism from Class Anthozoa. More preferably, the DNA has the sequence selected from the group consisting of SEQ ID Nos. 55, 57, 62, 64 and 66 and the fluorescent protein has the amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67.

In another embodiment of the present invention, there is provided a vector capable of expressing the DNA of the present invention in a recombinant cell comprising said DNA and regulatory elements necessary for expression of the DNA in the cell. Preferably, the DNA encodes a fluorescent protein having the amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67.

In still another embodiment of the present invention, there is provided a host cell transfected with a vector of the present invention, such that the host cell expresses a fluorescent protein. Preferably, the cell is selected from the group consisting of bacterial cells, mammalian cells, plant cells, insect cells and yeast cells. A representative example of bacterial cell is an *E. coli* cell.

The present invention is also directed to an isolated and purified fluorescent protein coded for by DNA selected from the group consisting of: (a) isolated DNA from an organism from Class Anthozoa which encodes a fluorescent protein; (b) isolated DNA which hybridizes to the isolated DNA of (a) and which encodes a fluorescent protein; and (c) isolated DNA differing from the isolated DNAs of (a) and (b) in

codon sequence due to the degeneracy of the genetic code, and which encodes a fluorescent protein. Preferably, the protein has the amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67.

5 The present invention is also directed to a DNA sequence encoding a fluorescent protein selected from the group consisting of: (a) an isolated DNA which encodes a fluorescent protein, wherein said DNA is from an organism from Class Anthozoa and wherein said organism does not exhibit bioluminescence; (b) an isolated DNA which 10 hybridizes to isolated DNA of (a) and which encodes a fluorescent protein; and (c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein. Preferably, the organism is from Sub-class Zoantharia, Order Corallimorpharia. More preferably, 15 the organism is from Family Discosomatidae, Genus Discosoma. Most particularly, the present invention is drawn to a novel fluorescent protein from *Discosoma* sp. "red", drFP583, or a fusion fluorescent protein drFP583/dmFP592.

The present invention is further directed to an amino acid 20 sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by means of hybridization, wherein the sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14. Preferably, such an oligonucleotide has a nucleotide sequence selected from the group 25 consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16.

Other and further aspects, features, and advantages of the present invention will be apparent from the following description of the presently preferred embodiments of the invention given for the purpose of disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the modified strategy of 3'-RACE used to isolate the target fragments. Sequences of the oligonucleotides used are shown in Table 2. Dp1 and Dp2 are the degenerate primers used in the first and second PCR, respectively (see Tables 3 and 4 for the sequences of degenerate primers). In the case of *Discosoma sp.* "red", the first degenerate primer used was NGH (SEQ ID No. 4), and the second degenerate primer used was GEGa (SEQ ID No. 6) or NFP (SEQ ID No. 13) or PVMb (SEQ ID No. 16).

Figure 2 shows the excitation and emission spectrum of the novel fluorescent protein from *Discosoma sp.* "red", drFP583.

Figure 3 shows separate detection of drFP583 and enhanced green fluorescent protein (EGFP) transiently expressed in HeLa cells. **Figure 3A** shows the image obtained for EGFP using Chroma filter set 31001. **Figure 3B** shows the image taken from the same field of view for drFP583 using filter set 31002. The images shown in Figures 3A and 3B were pseudocolored and overlaid in **Figure 3C**. **Figure 3D** shows phase contrast taken from the same field of view as that in Figure 3C.

Figure 4 shows separate detection of humanized drFP583, enhanced cyan fluorescent protein (ECFP) and enhanced yellow-green fluorescent protein (EYFP) in a triple-labeled HeLa cell. **Figure 4A** shows the image obtained from one cell co-expressing all three fluorescent proteins using Omega filter set XF 35 for DsRed1-Mito (drFP583-Mito, humanized drFP583). **Figure 4B** shows the image for EYFP-Tub using filter set XF 104. **Figure 4C** shows the image for ECFP-Nuc using filter set XF 114. The images shown in Figures 4A, 4B and 4C were pseudocolored and overlaid to show all three signals in one image in **Figure 4D**.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the term "GFP" refers to the basic green fluorescent protein from *Aequorea victoria*, including prior art versions of GFP engineered to provide greater fluorescence or fluoresce in different colors. The sequence of *Aequorea victoria* GFP (SEQ ID No. 54) has been disclosed in Prasher et al., *Gene* 111 (1992), 229-33.

As used herein, the term "EGFP" refers to mutant variant of GFP having two amino acid substitutions: F64L and S65T (Heim et al., 10 *Nature* 373 (1995), 663-664). The term "humanized" refers to changes made to the GFP nucleic acid sequence to optimize the codons for expression of the protein in human cells (Yang et al., *Nucleic Acids Research* 24 (1996), 4592-4593).

In accordance with the present invention there may be 15 employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual" (1982); "DNA Cloning: A Practical Approach," Volumes I and II (D.N. Glover ed. 20 1985); "Oligonucleotide Synthesis" (M.J. Gait ed. 1984); "Nucleic Acid Hybridization" (B.D. Hames & S.J. Higgins eds. (1985)); "Transcription and Translation" (B.D. Hames & S.J. Higgins eds. (1984)); "Animal Cell Culture" (R.I. Freshney, ed. (1986)); "Immobilized Cells and Enzymes" (IRL Press, (1986)); B. Perbal, "A Practical Guide To Molecular Cloning" 25 (1984).

A "vector" is a replicon, such as plasmid, phage or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "DNA molecule" refers to the polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in either single stranded form or a double-stranded helix. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes.

A DNA "coding sequence" is a DNA sequence which is transcribed and translated into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and synthetic DNA sequences. A polyadenylation signal and transcription termination sequence may be located 3' to the coding sequence.

As used herein, the term "hybridization" refers to the process of association of two nucleic acid strands to form an antiparallel duplex stabilized by means of hydrogen bonding between residues of the opposite nucleic acid strands.

The term "oligonucleotide" refers to a short (under 100 bases in length) nucleic acid molecule.

"DNA regulatory sequences", as used herein, are transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, and the like, that provide for and/or regulate expression of a coding sequence in a host cell.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive the various vectors of the present invention.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

A cell has been "transformed" or "transfected" by exogenous or heterologous DNA when such DNA has been introduced inside the cell. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell

"line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

A "heterologous" region of the DNA construct is an identifiable segment of DNA within a larger DNA molecule that is not found in association with the larger molecule in nature. Thus, when the heterologous region encodes a mammalian gene, the gene will usually be flanked by DNA that does not flank the mammalian genomic DNA in the genome of the source organism. In another example, heterologous DNA includes coding sequence in a construct where portions of genes from two different sources have been brought together so as to produce a fusion protein product. Allelic variations or naturally-occurring mutational events do not give rise to a heterologous region of DNA as defined herein.

As used herein, the term "reporter gene" refers to a coding sequence attached to heterologous promoter or enhancer elements and whose product may be assayed easily and quantifiably when the construct is introduced into tissues or cells.

The amino acids described herein are preferred to be in the "L" isomeric form. The amino acid sequences are given in one-letter code (A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; K: lysine; L: leucine; M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S: serine; T: threonine; V: valine; W: tryptophane; Y: tyrosine; X: any residue). NH₂ refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. In keeping with standard polypeptide nomenclature, *J Biol. Chem.*, 243 (1969), 3552-59 is used.

The present invention is directed to an isolated DNA selected from the group consisting of: (a) isolated DNA from an organism from the Class Anthozoa which encodes a fluorescent protein; (b) isolated DNA which hybridizes to isolated DNA of (a) and which encodes a fluorescent protein; and (c) isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to the degeneracy of the genetic code, and which encodes a fluorescent protein. Preferably, the DNA has the sequence selected from the group consisting of SEQ ID Nos. 55, 57, 62, 64 and 66, and the fluorescent protein has the amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67.

In another embodiment of the present invention, there is provided a vector capable of expressing the DNA of the present invention in a recombinant cell comprising said DNA and regulatory elements necessary for expression of the DNA in the cell. Specifically, the DNA encodes a fluorescent protein having the amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67.

In still another embodiment of the present invention, there is provided a host cell transfected with the vector of the present invention, which expresses a fluorescent protein of the present invention. Preferably, the cell is selected from the group consisting of bacterial cells, mammalian cells, plant cells and insect cells. A representative example of bacterial cell is an *E. coli* cell.

The present invention is also directed to a DNA sequence encoding a fluorescent protein selected from the group consisting of: (a) an isolated DNA which encodes a fluorescent protein, wherein said DNA is from an organism from Class Anthozoa and wherein said organism does not exhibit bioluminescence; (b) an isolated DNA which hybridizes to isolated DNA of (a) and which encodes a fluorescent

protein; and (c) an isolated DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein. Preferably, the organism is from Sub-class Zoantharia, Order Corallimorpharia. More preferably, 5 the organism is from Family Discosomatidae, Genus Discosoma. Most preferably, the organism is *Discosoma sp. "red"*.

The present invention is also directed to an isolated and purified fluorescent protein coded for by DNA selected from the group consisting of: (a) an isolated protein encoded by a DNA which encodes 10 a fluorescent protein wherein said DNA is from an organism from Class Anthozoa and wherein said organism does not exhibit bioluminescence; (b) an isolated protein encoded by a DNA which hybridizes to isolated DNA of (a); and (c) an isolated protein encoded by a DNA differing from the isolated DNAs of (a) and (b) in codon sequence due to 15 degeneracy of the genetic code. Preferably, the isolated and purified fluorescent protein is drFP583, or a fusion fluorescent protein drFP583/dmFP592.

The present invention is further directed to an amino acid sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by 20 means of hybridization, wherein the amino acid sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14. Preferably, such an oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16 and is used as 25 a primer in polymerase chain reaction. Alternatively, it can be used as a probe for hybridization screening of the cloned genomic or cDNA library.

The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion.

5

EXAMPLE 1

Biological Material

Novel fluorescent proteins were identified from several 10 genera of Anthozoa which do not exhibit any bioluminescence but have fluorescent color as observed under usual white light or ultraviolet light. Six species were chosen (see Table 1).

TABLE 1Anthozoa Species Used in This Study

Species	Area of Origination	Fluorescent Color
Anemonia majano	Western Pacific	bright green tentacle tips
Clavularia sp.	Western Pacific	bright green tentacles and oral disk
Zoanthus sp.	Western Pacific	green-yellow tentacles and oral disk
Discosoma sp. "red"	Western Pacific	orange-red spots oral disk
Discosoma striata	Western Pacific	blue-green stripes on oral disk
Discosoma sp. "magenta"	Western Pacific	faintly purple oral disk
Discosoma sp. "green"	Western Pacific	green spots on oral disk
Anemonia sulcata	Mediterranean	purple tentacle tips

5

EXAMPLE 2cDNA Preparation

Total RNA was isolated from the species of interest
 10 according to the protocol of Chomczynski and Sacchi (Chomczynski P.,

et al., *Anal. Biochem.* 162 (1987), 156-159). First-strand cDNA was synthetized starting with 1-3 µg of total RNA using SMART PCR cDNA synthesis kit (CLONTECH) according to the provided protocol with the only alteration being that the "cDNA synthesis primer" provided in the 5 kit was replaced by the primer TN3 (5'- CGCAGTCGACCG(T)₁₃, SEQ ID No. 1) (Table 2). Amplified cDNA samples were then prepared as described in the protocol provided except the two primers used for PCR were the TS primer (5'-AAGCAGTGGTATCAACGCAGAGT, SEQ ID No. 2) (Table 2) and the TN3 primer (Table 2), both in 0.1 µM concentration. 10 Twenty to twenty-five PCR cycles were performed to amplify a cDNA sample. The amplified cDNA was diluted 20-fold in water and 1 µl of this dilution was used in subsequent procedures.

TABLE 2Oligos Used in cDNA Synthesis and RACE

5 TN3: 5'-CGCAGTCGACCG(T)₁₃
(SEQ ID No. 1)

T7-TN3: 5'-GTAATACGACTCACTATAAGGCCGCAGTCGACCG(T)₁₃
(SEQ ID No. 17)

10 TS-primer: 5'-AAGCAGTGGTATCAACGCAGAGT
(SEQ ID No. 2)

15 T7-TS:
5'-GTAATACGACTCACTATAAGGCAGCAGTGGTATCAACGCAGAGT
(SEQ ID No. 18)

T7: 5'-GTAATACGACTCACTATAAGGC
(SEQ ID No. 19)

20 TS-oligo 5'-AAGCAGTGGTATCAACGCAGAGTACGCrGrGrG
(SEQ ID No. 53)

EXAMPLE 3

25

Oligo Design

To isolate fragments of novel fluorescent protein cDNAs, PCR using degenerate primers was performed. Degenerate primers were designed to match the sequence of the mRNAs in regions that were predicted to be the most invariant in the family of fluorescent proteins. Four such stretches were chosen (Table 3) and variants of degenerate primers were designed. All such primers were directed to

the 3'-end of mRNA. All oligos were gel-purified before use. Table 2 shows the oligos used in cDNA synthesis and RACE.

TABLE 3

5

Key Amino Acid Stretches and Corresponding Degenerate Primers Used for Isolation of Fluorescent Proteins

Stretch Position according to A. victoria GFP (7)	Amino Acid Sequence of the Key Stretch	Degenerated Primer Name and Sequence
20-25	GXVN ^G H (SEQ ID No. 3)	NGH: 5'- GA(C,T) GGC TGC GT(A,T,G,C) AA(T,C) GG(A,T,G) CA (SEQ ID No. 4)
31-35	GEGEG (SEQ ID No. 5) GEGNG (SEQ ID No. 8)	GEGA: 5'- GTT ACA GGT GA(A,G) GG(A,C) GA(A,G) GG (SEQ ID No. 6) GEGb: 5'- GTT ACA GGT GA(A,G) GG(T,G) GA(A,G) GG (SEQ ID No. 7) GN ^G a: 5'- GTT ACA GGT GA(A,G) GG(A,C) AA(C,T) GG (SEQ ID No. 9) GN ^G b: 5'- GTT ACA GGT GA(A,G) GG(T,G) AA(C,T) GG (SEQ ID No. 10)
127-131	GMNFP (SEQ ID No. 11) GVNFP (SEQ ID No. 12)	NFP: 5' TTC CA(C,T) GGT (G,A)TG AA(C,T) TT(C,T) CC (SEQ ID NO. 13)
134-137	GPVM (SEQ ID No. 14)	PV ^G a: 5' CCT GCC (G,A)A(C,T) GGT CC(A,T,G,C) GT(A,C) ATG (SEQ ID NO. 15) PV ^G b: 5' CCT GCC (G,A)A(C,T) GGT CC(A,T,G,C) GT(G,T) ATG (SEQ ID NO. 16)

EXAMPLE 4**Isolation of 3'-cDNA Fragments of nFPs**

The modified strategy of 3'-RACE was used to isolate the target fragments (see Figure 1). The RACE strategy involved two consecutive PCR steps. The first PCR step involved a first degenerate primer (Table 4) and the T7-TN3 primer (SEQ ID No. 17) which has a 3' portion identical to the TN3 primer used for cDNA synthesis (for sequence of T7-TN3, Table 2). The reason for substituting the longer T7-TN3 primer in this PCR step was that background amplification which occurred when using the shorter TN3 primer was suppressed effectively, particularly when the T7-TN3 primer was used at a low concentration (0.1 μ M) (Frohman et al., (1998) *PNAS USA*, 85, 8998-9002). The second PCR step involved the TN3 primer (SEQ ID No. 1, Table 2) and a second degenerate primer (Table 4).

TABLE 4

Combinations of Degenerate Primers for First and Second PCR Resulting in Specific Amplification of 3'-Fragments of nFP cDNA

Species	First Degenerate Primer	Second Degenerate Primer
<i>Anemonia majano</i>	NGH (SEQ ID No. 4)	GNGb (SEQ ID No. 10)
<i>Clavularia</i> sp.	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6)
<i>Zoanthus</i> sp.	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6)
<i>Discosoma</i> sp. "red"	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6), NFP (SEQ ID No. 13) or PVMb (SEQ ID No. 16)
<i>Discosoma striata</i>	NGH (SEQ ID No. 4)	NFP (SEQ ID No. 13)
<i>Anemonia sulcata</i>	NGH (SEQ ID No. 4)	GEGa (SEQ ID No. 6) or NFP (SEQ ID No. 13)

5

The first PCR reaction was performed as follows: 1 µl of 20-fold dilution of the amplified cDNA sample was added into the reaction mixture containing 1X Advantage KlenTaq Polymerase Mix with provided buffer (CLONTECH), 200 µM dNTPs, 0.3 µM of first degenerate

primer (Table 4) and 0.1 μ M of T7-TN3 (SEQ ID No. 17) primer in a total volume of 20 μ l. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 1 cycle for 95°C, 10 sec.; 55°C, 1 min.; 72°C, 40 sec; 24 cycles for 95°C, 10 sec.; 62°C, 30 sec.; 72°C, 40 sec. The reaction was then diluted 20-fold in water and 1 μ l of this dilution was added to a second PCR reaction, which contained 1X Advantage KlenTaq Polymerase Mix with the buffer provided by the manufacturer (CLONTECH), 200 μ M dNTPs, 0.3 μ M of the second degenerate primer (Table 4) and 0.1 μ M of TN3 primer. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 1 cycle for 95°C, 10 sec.; 55°C (for GEG/GNG or PVM) or 52°C (for NFP), 1 min.; 72°C, 40 sec; 13 cycles for 95°C, 10sec.; 62°C (for GEG/GNG or PVM) or 58°C (for NFP), 30 sec.; 72°C, 40 sec. The product of PCR was cloned into PCR-Script vector (Stratagene) according to the manufacturer's protocol.

Different combinations of degenerate primers were tried in the first and second PCR reactions on the DNA from each species until a combination of primers was found that resulted in specific amplification--meaning that a pronounced band of expected size (about 650-800 bp for NGH and GEG/GNG and 350-500 bp for NFP and PVM--sometimes accompanied by a few minor bands) was detected on agarose gel after two PCR reactions. The primer combinations of choice for different species of the Class Anthozoa are listed in Table 4. Some other primer combinations also resulted in amplification of fragments of correct size, but the sequence of these fragments showed no homology to the other fluorescent proteins identified or to *Aequorea victoria* GFP.

EXAMPLE 5**Obtaining Full-Length cDNA Copies**

Upon sequencing the obtained 3'-fragments of novel fluorescent protein cDNAs, two nested 5'-directed primers were synthesized for cDNA (Table 5), and the 5' ends of the cDNAs were then amplified using two consecutive PCRs. In the next PCR reaction, the novel approach of "step-out PCR" was used to suppress background amplification. The step-out reaction mixture contained 1x Advantage KlenTaq Polymerase Mix using buffer provided by the manufacturer (CLONTECH), 200 µM dNTPs, 0.2 µM of the first gene-specific primer (see Table 5), 0.02 µM of the T7-TS primer (SEQ ID No. 18), 0.1 µM of T7 primer (SEQ ID No. 19) and 1 µl of the 20-fold dilution of the amplified cDNA sample in a total volume of 20 µl. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 23-27 cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of amplification was diluted 50-fold in water and one µl of this dilution was added to the second (nested) PCR. The reaction contained 1X Advantage KlenTaq Polymerase Mix with provided buffer (CLONTECH), 200 µM dNTPs, 0.2 µM of the second gene-specific primer and 0.1 µM of TS primer (SEQ ID No. 2) in a total volume of 20 µl. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 12 cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of amplification was then cloned into pAtlas vector (CLONTECH) according to the manufacturer's protocol.

TABLE 5Gene-Specific Primers Used for 5'-RACE

Species	First Primer	Second (Nested) Primer
<i>Anemonia majano</i>	5'-GAAATAGTCAGGCATACTGGT (SEQ ID No. 20)	5'-GTCAGGCATAC TGGTAGGAT (SEQ ID No. 21)
<i>Clavularia</i> sp.	5'-CTTGAAATAGTCTGCTATATC (SEQ ID No. 22)	5'-TCTGCTATATC GTCTGGGT (SEQ ID No. 23)
<i>Zoanthus</i> sp.	5'- GTTCTTGAAATAGTCTACTATGT (SEQ ID No. 24)	5'-GTCTACTATGTCTT GAGGAT (SEQ ID No. 25)
<i>Discosoma</i> sp. "red"	5'-CAAGCAAATGGCAAAGGTC (SEQ ID No. 26)	5'-CGGTATTGTGGCC TTCGTA (SEQ ID No. 27)
<i>Discosoma</i> <i>striata</i>	5'-TTGTCTTCTTCTGCACAAC (SEQ ID No. 28)	5'-CTGCACAAACGG GTCCAT (SEQ ID No. 29)
<i>Anemonia</i> <i>sulcata</i>	5'-CCTCTATCTTCATTCCTGC (SEQ ID No. 30)	5'-TATCTTCATTCCT GCGTAC (SEQ ID No. 31)
<i>Discosoma</i> sp. "magenta"	5'-TTCAGCACCCCATCACGAG (SEQ ID No. 32)	5'-ACGCTCAGAGCTG GGTTCC (SEQ ID No. 33)
<i>Discosoma</i> sp. "green"	5'-CCCTCAGCAATCCATCACGTTC (SEQ ID No. 34)	5'-ATTATCTCAGTGGA TGGTTC (SEQ ID No. 35)

EXAMPLE 6Expression of nFPs in *E.coli*

To prepare a DNA construct for novel fluorescent protein expression, two primers were synthesized for each cDNA: a 5'-directed "downstream" primer with the annealing site located in the 3'-UTR of the cDNA and a 3'-directed "upstream" primer corresponding to the site of translation start site (not including the first ATG codon) (Table 6). Primers with SEQ ID Nos. 43 and 44 were the primers used to prepare the dr583 DNA. Both primers had 5'-heels coding for a site for a restriction endonuclease; in addition, the upstream primer was designed so as to allow the cloning of the PCR product into the pQE30 vector (Qiagen) in such a way that resulted in the fusion of reading frames of the vector-encoded 6xHis-tag and nFP. The PCR was performed as follows: 1 µl of the 20-fold dilution of the amplified cDNA sample was added to a mixture containing 1x Advantage KlenTaq Polymerase Mix with buffer provided by the manufacturer (CLONTECH), 200 µM dNTPs, 0.2 µM of upstream primer and 0.2 µM of downstream primer, in a final total volume of 20 µl. The cycling profile was (Hybaid OmniGene Thermocycler, tube control mode): 23-27 cycles for 95°C, 10 sec.; 60°C, 30 sec.; 72°C, 40 sec. The product of this amplification step was purified by phenol-chlorophorm extraction and ethanol precipitation and then cloned into pQE30 vector using restriction endonucleases corresponding to the primers' sequence according to standard protocols.

All plasmids were amplified in XL-1 blue *E. coli* and purified by plasmid DNA miniprep kits (CLONTECH). The recombinant clones were selected by colony color, and grown in 3 ml of LB medium (supplemented with 100 µg/ml of ampicillin) at 37°C overnight. 100 µl

of the overnight culture was transferred into 200 ml of fresh LB medium containing 100 µg/ml of ampicillin and grown at 37°C, 200 rpm up to OD₆₀₀ 0.6-0.7. 1 mM IPTG was then added to the culture and incubation was allowed to proceed at 37°C for another 16 hours. The
5 cells were harvested and recombinant protein, which incorporated 6x His tags on the N-terminus, was purified using TALON™ metal-affinity resin according to the manufacturer's protocol (CLONTECH).

TABLE 6

Primers Used to Obtain Full Coding Region of nFPs for Cloning into Expression Construct

Species	Upstream Primer	Downstream Primer
Anemonia majano	5' -acatggatccgctttcaaaca agtttac (SEQ ID No. 36) BamHI	5'-tagtactcgagcttattcgta tttcagtgaaatc (SEQ ID No. 37) XhoI
Clavularia sp.	L: 5'-acatggatccaacatTTTGA gaaacg (SEQ ID No. 38) BamHI S: 5'-acatggatccaagctctaacc accatg (SEQ ID No. 39) BamHI	5'-tagtactcgagcaacacaa accctcagacaa (SEQ ID No. 40) XhoI
Zoanthus sp.	5'- acatggatccgctcagtcaaag cacgg (SEQ ID No. 41) BamHI	5'-tagtactcgagggttggaaactacat tcttatca (SEQ ID No. 42) XhoI
Discosoma sp. "red"	5'- acatggatccaggtttcaagaat gttatac (SEQ ID No. 43) BamHI	5'-tagtactcgaggaggccaagttc agccta (SEQ ID No. 44) XhoI
Discosoma striata	5'- acatggatccagttgttccaagagtgt (SEQ ID No. 45) BamHI	5'-tagcgagcttatcatgcctc gtcacct (SEQ ID No. 46) SacI
Anemonia sulcata	5'- acatggatccgcttcttttaaagaagact (SEQ ID No. 47) BamHI	5'-tagtactcgagtcttgggagc ggcttg (SEQ ID No. 48) XhoI
Discosoma sp. "magenta"	5'- acatggatccagttgttccaagaatgtat (SEQ ID No. 49) BamHI	5'-tagtactcgaggccattacg ctaatac (SEQ ID No. 50) XhoI
Discosoma sp. "green"	5'-acatggatccagtgcaactaaagaagaatg (SEQ ID No. 51)	5'-tagtactcgagattcggttaat gccttg (SEQ ID No. 52)

EXAMPLE 7**Novel Fluorescent Protein drFP583 and cDNA Encoding drFP583**

One of the full-length cDNAs encoding fluorescent proteins found is described herein (drFP583). The nucleic acid sequence and deduced amino acid sequence are SEQ ID Nos. 55 and 56, respectively. The spectral properties of drFP583 is listed in Table 7, and the emission and excitation spectra for the drFP583 is shown in Figure 2.

10

TABLE 7**Spectral Properties of the Isolated drFP583**

15	Species:	Discosoma sp. "red"	Max. Extinction Coefficient:	22,500
	nFP Name:	drFP583	Quantum Yield	0.23
20	Absorbance Max. (nm):	558	Relative Brightness:*	0.24
	Emission Max. (nm):	583		

*relative brightness is extinction coefficient multiplied by quantum yield divided by the same value for *A. victoria* GFP.

25

EXAMPLE 8**Expression of drFP583 in Mammalian Cells**

30 HeLa cells were transfected either with plasmid pDsRed1-N1 (vector containing the DNA encoding drFP583) or plasmid pEGFP-C1

(encoding EGFP from *Aequoria victoria*). Immediately after the transfection, cells were mixed and plated on coverslips. Cells were incubated for 48 hours at 37°C then fixed in 3.7 % formaldehyde. Cells were mounted in mounting medium and observed by fluorescence microscopy. Images were taken from the same field of view with Chroma filter set 31001 for EGFP (Figure 3A), and filter set 31002 for drFP583 (Figure 3B) using a cooled CCD camera (Roper Scientific) and MetaMorph software (Universal Imaging). The images (Figures 3A and 3B) were pseudocolored and overlayed in Figure 3C. Phase contrast was taken from the same field of view and overlayed in Figure 3D.

EXAMPLE 9**Generation of Humanized drFP583**

Since mammalian expression is a very popular tool, human favored codon version is needed for better expression in mammalian cells. Humanized drFP583 was therefore generated by changing wild type drFP583 nucleotide sequence to optimize the codons for expression of the fluorescent protein. The nucleotide sequence of the humanized drFP583 is shown in SEQ ID No. 57.

10

EXAMPLE 10**Expression of Humanized drFP583 in Mammalian Cells**

HeLa cells were transiently co-transfected with plasmids pECFP-Nuc, pEYFP-Tub and pDsRed1-Mito (humanized drFP583). After transfection, cells were incubated for 48 hours at 37 °C then fixed in 3.7 % formaldehyde. Cells were mounted in mounting medium and observed by fluorescence microscopy. Images were taken of one cell co-expressing all three fluorescent proteins with Omega filter set XF 35 for DsRed1-Mito (Figure 4A), XF 104 for EYFP-Tub (Figure 4B) and XF 114 for ECFP-Nuc (Figure 4C) using a cooled CCD camera (Roper Scientific) and MetaMorph software (Universal Imaging). Individual images were pseudocolored and overlayed to show all three signals in one image (Figure 4D). Protein DsRed1-Mito localizes to mitochondria, EYFP-Tub localizes to the microtubular network, and ECFP-Nuc localizes to the nucleus.

As a conclusion, drFP583 does emit to a low extent also in the cyan (ECFP), green (EGFP) and yellow-green (EYFP) emission channels

(filter sets). High expression levels or highly concentrated protein in intracellular structures can therefore result in high signal intensities that will give some bleedthrough in the other emission wavelengths. The bleedthrough is small and should not affect multiple labeling 5 recording in most cases.

EXAMPLE 11

10 Mutants of Humanized drFP583

Mutants of humanized drFP583 were generated using error prone PCR technique (Clontech). Mutations occurred at amino acids 42, 71, 105, 120, 161 and 197 (numbering starting from the first Methionine). Table 8 lists the mutants that were generated and their 15 properties.

TABLE 8Mutants of Humanized drFP583

5

Mutant	Mutations	Properties
E5	V105A, S197T	Overnight in <i>E. coli</i> emitting green fluorescence; <i>in vitro</i> maturing to red over at 28h 37°C on 80% (retains 20% green peak); folding faster than wild type drFP583 (~28h)
E8	N42H	always two peaks green & red (~1:1) folding faster than E5 (~8h)
E5up	V105A	red from the beginning; folding faster than E5 (~12h)
E5down	S197T	phenotype is similar to E5
E57	V105A, I161,S197A	like E5 but folding faster (~8-10h); ~5% of green peak left at the end
AG4	V71M, V105A, S197T	bright green, no red at all; fast folding (~ 16h)
AG4H	V71M, V105A, Y120H, S197T	like AG4 but twice brighter

EXAMPLE 12**Characterization and Applications of E5 Mutant**

5 E5 (V105A, S197T) changes its fluorescence from green to red over time both *in vitro* and *in vivo*, in *E. coli* and in mammalian cells. Also, E5 folds faster than wild type drFP583 both in *E. coli* and mammalian cells (Table 8).

10 Since it allows the "two color" reporting mode for monitoring of the promoter activity, i.e., for both active or shutdown state of the promoter, there is a separate color, serving as an indicator of that state, E5 can be used as a transcriptional reporter. Different from "one color" mode, "two color" mode has a measurable signal (color) present for both states of the process as opposed to "one 15 color" mode (e.g. destabilized GFP) wherein the absence of the color is an indicator of the second state. Namely, newly produced E5 protein fluoresces in green, indicating on-going promoter activity. Over time, the protein will mature, acquiring the red fluorescence. So if the promoter is no longer active, all the protein will eventually mature, 20 resulting in the dominant red fluorescence. In case the promoter is still active both red and green fluorescence will be readily detected. Thus E5 as a "two color" reporter allows study of gene expression similar to destabilized GFP, but with permanent "signature" of past gene activity in the cells, tissues or entire organism. For example, at the 25 tissue level, E5 may help to distinguish the stem cells from differentiated cells. Providing the promoter is only active in the stem cells, the E5 reporter will label the stem cell population in green and red, the progenitor cells would be labeled predominantly in red, the terminally differentiated will not fluoresce (due to the titration out of 30 protein during cell division).

E5 could be used for spatial and temporal visualization of newly synthesized vs. accumulated fusion proteins. That is, E5 could function like a fusion tag. Possible applications envisaged at different organizational levels. At the cellular level, E5 may help to visualize and 5 distinguish the newly synthesized proteins in various compartments such as outer membrane, microvillae, ER, Golgi, mitochondria, nuclei, various components of cellular matrix and focal adhesion complexes. At the tissue level, E5 may be helpful in visualizing newly formed vs. preexisting structures e.g. membrane junctions, components of 10 extracellular matrix.

One of the most fascinating applications of E5 seems to be in the study of the mother-daughter relationship during the cell division and migration. A wide horizon is opening in the fields of development as well as in the adult organisms to study the cell 15 migration and differentiation. Allowing visualization of the expression "history", E5 can help to distinguishing between the mother cells where the promoter is actually active vs. the daughter cells containing the accumulated protein but not producing fresh protein anymore. It would enable the study of the cell fates during development and organ 20 remodeling, thus distinguishing between cell migration and cell expansion or differentiation.

In conclusion, E5 is basically applicable to any situation where GFP was previously used. Main advantage is that E5 can track down "the history" of promoter activity or protein localization in cells 25 or tissues. With a better protein stability than GFP, E5 allows longer analysis window (wild type drFP583 is stable for at least 4 weeks in Xenopus, while EGFP starts to faint after two weeks).

EXAMPLE 13

Characterization and Applications of E8 Mutant

E8 (N42H) has two fluorescence maximums, green and red,
5 at all times and it folds much faster than drFP583 (Table 8).

Since it detects both green and red fluorescence simultaneously, E8 may be useful for studying processes related to blood circulation and proteins/cells trafficking. Blood absorbs the green fluorescence; thus only the red fluorescence will be visible while
10 the protein is trafficking in the blood. Both green and red fluorescence could be detected outside the bloodstream making the whole process easy to visualize and record. Monitoring both red and green fluorescence simultaneously may also help to reduce the background fluorescence problems for some tissues or cells.

15

EXAMPLE 14

Generation of drFP583/dmFP592 Hybrid Using Shuffling Procedure

20 Non-humanized wild type coding region fragments from drFP583 and dmFP592 were amplified by PCR (22 cycles, 95°C, 15 sec., 68°C 1 min 20 sec.) using 1 ng of corresponding bacterial expression plasmids (pQE-30 derivatives with drFP583 and dmFP592 inserts, respectively) as templates. Oligonucleotides A (ACATGG
25 ATCCAGGTCTTCCAAGAATGTTATC, SEQ ID No. 58), B (TAGTACTCG AGCCAAGTTCAGCCTTA, SEQ ID No. 59), C (ACATGGATCCAG TTGTTC CAAGAATGTGAT, SEQ ID No. 60), and D (TAGTACTCGAGGCCATTA CCGCTAAC, SEQ ID No. 61) were used as primers for amplifying these fragments in a concentration of 0.2 mM.

The PCR products were then purified by QIAquick PCR Purification Kit (QIAGEN). Afterwards, the purified fragments drFP583 and dmFP592 (300-500 ng each) were digested with restriction endonucleases EcoRI, HindIII and DraI (10 U each) simultaneously.

5 Reactions were performed in BamHI restriction buffer (NEB) supplemented BSA for 3h at 37°C. Total reaction volume was 30 ml. Upon completion, the resulted restriction fragments from each restriction reaction were separated by electrophoresis in agarose gels (1.5%), cut from gel and purified by QIAquick Gel Purification Kit

10 (QIAGEN). The resulting sets of the purified restriction fragments from both drFP583 and dmFP592 were combined together and 50 ng of them were put into ligation mix (1X T4 DNA ligation buffer, 400 NEB U of T4 DNA ligase) in total volume of 30 ml. The ligation was performed for 3h at room temperature and stopped by heating at 70°C within 20 min.

15 The ligation mixture was then diluted by water ten-folds, and 1 ml of the dilution was preformed for PCR reaction (20 cycles, 95°C, 15 sec. 68°C 1min 20 sec) as template. Four oligonucleotides A, B, C, and D (SEQ ID Nos. 58-61, respectively) were used simultaneously as primers for amplifying these fragments in a concentration of 0.1 mM each. After electrophoresis in an agarose gel (1.5%), the target fragment was purified by QIAquick Gel Purification Kit (QIAGEN) and digested with restriction endonucleases BamHI and XhoI (30-50 U each) simultaneously in BamH I restriction buffer (NEB) supplemented with BSA for 3h at 37°C. After purification, the resulting fragment was

20 cloned in pQE-30 plasmid linearized by BamHI and SalI. Ligation reaction was performed in 1X T4 DNA ligation buffer and 400 NEB U of T4 DNA ligase with a total volume of 20 ml for overnight at 16°C. After transformation of *E.coli* cells by 1/5 of the ligation volume and incubation on LB/1% agar plates which were supplemented by 100

25

mg/ml Ampacilin and 0.1 mM IPTG at 37°C for overnight, the resulting *E.coli* colonies were screened visually under fluorescent microscope using rhodamine filter set. The brightest red colonies were picked up and placed in 200 ml LB medium with 100 mg/ml of Ampacilin. At 5 OD₆₀₀=0.6, the *E.coli* culture was induced by IPTG (final concentration was 1 mM) and the fermentation continued for overnight. Purification of recombinant protein containing N-terminus 6Xhis tag was performed using TALON metal-affinity resin according to manufacturer's protocol.

10

EXAMPLE 15

Spectral Properties of drFP583/dmFP592 Hybrid

The emission and excitation spectra for drFP583/dmFP592 hybrid protein are basically the same as for dmFP592. Table 9 lists the spectral properties of drFP583/dmFP592 hybrid protein.

TABLE 9

20 **Spectral Properties of drFP583/dmFP592 Hybrid**

nFP Name	Absorbance Maximum nm	Emission Maximum nm	Maximum Extinction Coeff.	Relative Quantum Yield*	Relative Brightness**
drFP583/ dmFP592	573	592	35,000	0.24	0.3

*relative quantum yield was determined as compared to the quantum yield of *A. victoria* GFP.

25 **relative brightness is extinction coefficient multiplied by quantum yield divided by the same value for *A. victoria* GFP.

EXAMPLE 16

Humanized drFP583/dmFP592 Hybrid and Mutants

drFP583/dmFP592 hybrid was humanized (SEQ ID No. 62 for 5 nucleotide sequence and SEQ ID No. 63 for deduced amino acid sequence). Further, two mutants were generated based on the humanized drFP583/dmFP592, i.e., drFP583/dmFP592-2G and drFP583/dmFP592-Q3. drFP583/dmFP592-2G contains two substitutions, K15Q and T217S (SEQ ID Nos. 64 and 65 for nucleotide 10 and deduced amino acid sequences, respectively). This mutant was derived from the humanized drFP583/dmFP592 hybrid gene by random mutagenesis using Diversity PCR Mutagenesis Kit (Clontech) according to the corresponding protocol. drFP583/dmFP592-Q3 contains three substitutions, K15Q and K83M and T217S (SEQ ID Nos. 66 and 67 for 15 nucleotide and deduced amino acid sequences, respectively). drFP583/dmFP592-Q3 mutant was derived from drFP583/dmFP592-2G mutant by random mutagenesis using Diversity PCR Mutagenesis Kit (Clontech) according to the corresponding protocol.

drFP583/dmFP592-2G has similar brightness and folding 20 rate as for non-humanized drFP583/dmFP592 hybrid. While drFP583/dmFP592-Q3 could be seen in *E.coli* cells as more dark red than parental variant, i.e., drFP583/dmFP592-2G, and the purified protein solution has purple color. drFP583/dmFP592-Q3 has the emission maximum of 616 nm and excitation maximum of 580 nm.

EXAMPLE 17**Possible Applications of Hybrid Mutants**

Similar to fluorescent protein drFP583 or dmFP592,
5 drFP583/dmFP592-Q3 could be used as a tool for investigation of
protein expression, transport and protein interactions *in vivo*,
monitoring of promoter activity, and as a transcription reporter or
fusion tag. Besides, drFP583/dmFP592-Q3 could be chosen as the most
convenient partner to one of the existing green fluorescent protein
10 variants in two/triple color labeling assays for simultaneous detection
of expression of two or more proteins *in vivo* due to its strongly red-
shifted position of emission maximum and practical absence of
excitation in green part of spectrum except any spectral overlapping
and background fluorescence.

15 The method of generating drFP583/dmFP592 hybrid can
have a general utility for generating hybrid genes (i.e., genes containing
parts of different genes in various combinations) with improved
fluorescent characteristics.

20 Additionally, drFP583/dmFP592-Q3 is the first red-shifted
mutant, which demonstrates that spectral-shifted mutant could be
obtained by random mutagenesis.

Any patents or publications mentioned in this specification
are indicative of the levels of those skilled in the art to which the
invention pertains. These patents and publications are incorporated by
25 reference to the same extent as if each individual publication was
specifically and individually indicated to be incorporated by reference.

One skilled in the art will appreciate readily that the present
invention is adapted to carry out the objects and obtain the ends and
advantages mentioned, as well as those objects and ends inherent

therein. The present examples, along with the methods, procedures, treatments, molecules, and specific compounds described herein, are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes 5 to the methods and compounds, and other uses, will occur to those skilled in the art and are encompassed within the spirit of the invention as defined by the scope of the claims.

WHAT IS CLAIMED IS:

1. A DNA sequence encoding a fluorescent protein selected from the group consisting of:

5 (a) an isolated DNA which encodes a fluorescent protein, wherein said DNA is from an organism from a Class Anthozoa and wherein said organism does not exhibit bioluminescence;

(b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

10 (c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

15 2. The DNA sequence of claim 1, wherein said organism is from Sub-class Zoantharia.

20 3. The DNA sequence of claim 2, wherein said organism is from Order Corallimorpharia.

4. The DNA sequence of claim 3, wherein said organism is from Family Discosomatidae.

25

5. The DNA sequence of claim 4, wherein said organism is from Genus Discosoma.

6. A DNA sequence encoding a fluorescent protein selected from the group consisting of:

(a) an isolated DNA which encodes a fluorescent protein, wherein said DNA has a sequence selected from the group consisting of

5 SEQ ID Nos. 55, 57, 62, 64 and 66;

(b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic

10 code, and which encodes a fluorescent protein.

7. The DNA sequence of claim 6, wherein said DNA encodes a fluorescent protein having an amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67.

15

8. The DNA sequence of claim 6, wherein said DNA is selected from the group consisting of non-humanized and humanized DNA.

20

9. A vector capable of expressing the DNA sequence of claim 1 in a recombinant cell, wherein said vector comprising said DNA and regulatory elements necessary for expression of the DNA in the 25 cell.

10. The vector of claim 9, wherein said DNA encodes a fluorescent protein having the amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67.

5

11. The vector of claim 9, wherein said DNA is selected from the group consisting of non-humanized and humanized DNA.

10 12. A host cell transfected with the vector of claim 9, wherein said cell is capable of expressing a fluorescent protein.

15 13. The host cell of claim 12, wherein said cell is selected from the group consisting of bacterial cells, mammalian cells, plant cell, yeast and insect cells.

20 14. The host cell of claim 13, wherein said bacterial cell is an *E. coli* cell.

15. An isolated and purified fluorescent protein coded for by DNA selected from the group consisting of:

25 (a) an isolated DNA which encodes a fluorescent protein from an organism from Class Anthozoa, wherein said organism does not exhibit bioluminescence;

(b) an isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) an isolated DNA differing from the isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

5

16. The isolated and purified fluorescent protein of claim 15, wherein said organism is from Sub-class Zoantharia.

10

17. The isolated and purified fluorescent protein of claim 16, wherein said organism is from Order Corallimorpharia.

15

18. The isolated and purified fluorescent protein of claim 17, wherein said organism is from Family Discosomatidae.

20

19. The isolated and purified fluorescent protein of claim 18, wherein said organism is from Genus Discosoma.

20. An isolated and purified fluorescent protein coded for by DNA selected from the group consisting of:

(a) isolated DNA which encodes a fluorescent protein having an amino acid sequence selected from the group consisting of SEQ ID Nos. 56, 63, 65 and 67;

(b) isolated DNA which hybridizes to isolated DNA of (a) above and which encodes a fluorescent protein; and

(c) isolated DNA differing from said isolated DNAs of (a) and (b) above in codon sequence due to degeneracy of the genetic code and which encodes a fluorescent protein.

5

21. The isolated and purified fluorescent protein of claim 20, wherein said protein is drFP583.

10 22. The isolated and purified fluorescent protein of claim 20, wherein said protein is drFP583/dmFP592 hybrid protein.

15 23. An amino acid sequence which can be used as a basis for designing an oligonucleotide probe for identification of a DNA encoding a fluorescent protein by means of hybridization, wherein said sequence is selected from the group consisting of SEQ ID Nos. 3, 5, 8, 11, 12, 14.

20 24. The amino acid sequence of claim 23, wherein said oligonucleotide has a nucleotide sequence selected from the group consisting of SEQ ID Nos. 4, 6, 7, 9, 10, 13, 15, 16.



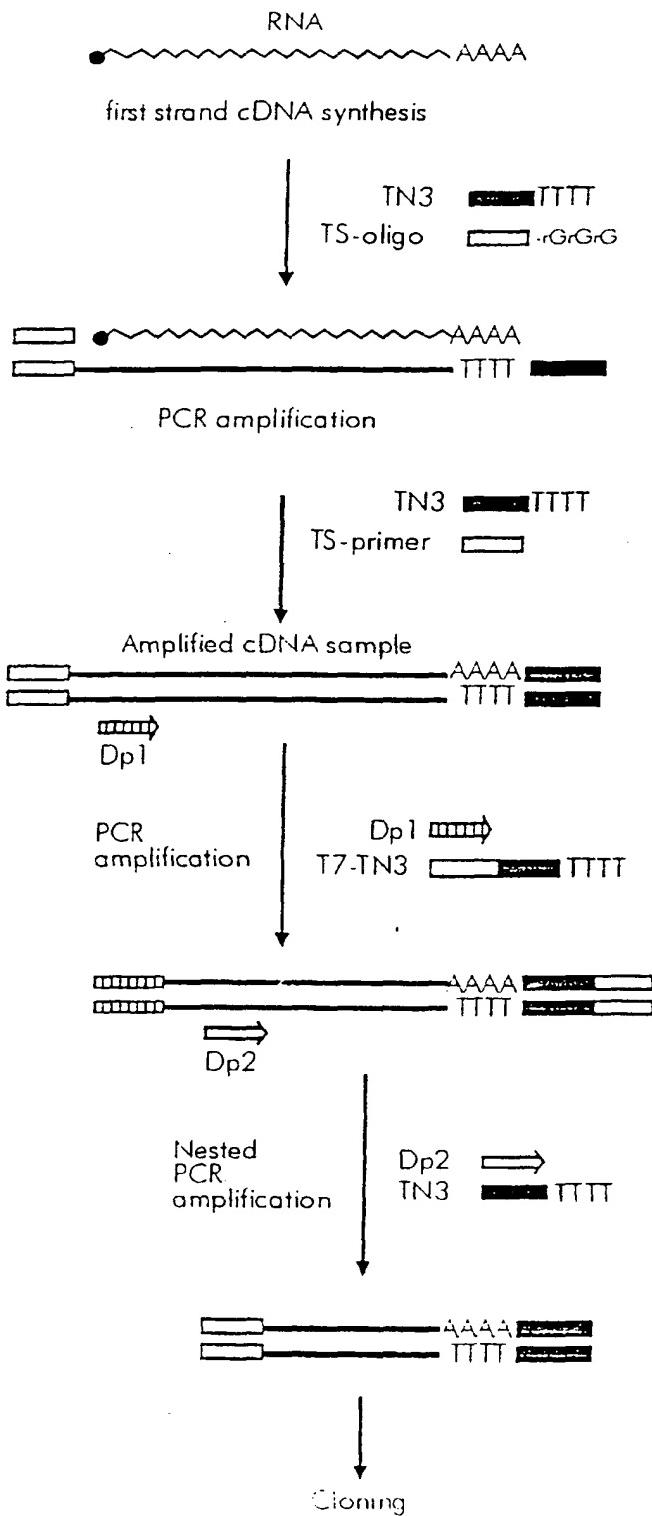
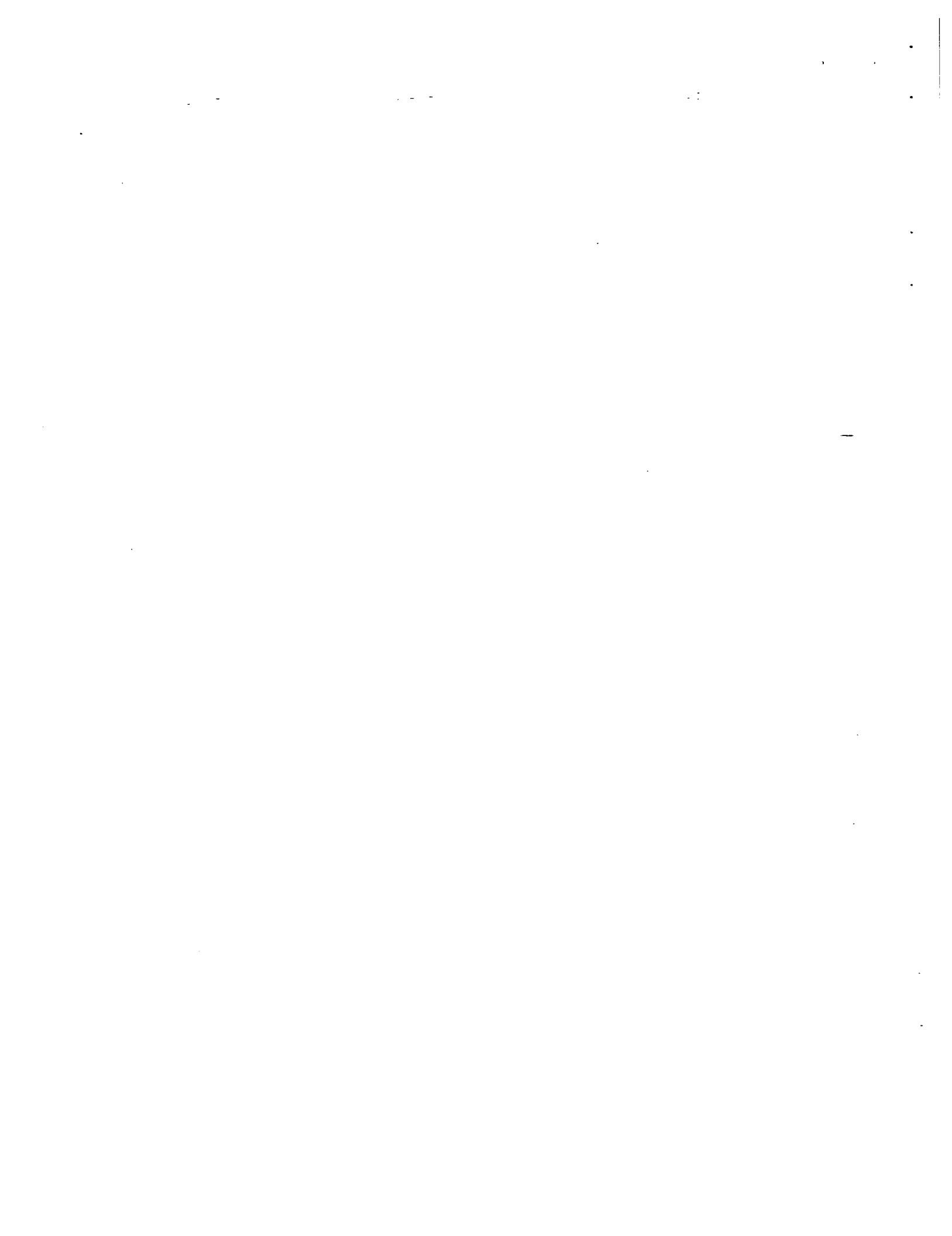


Figure 1



2/10

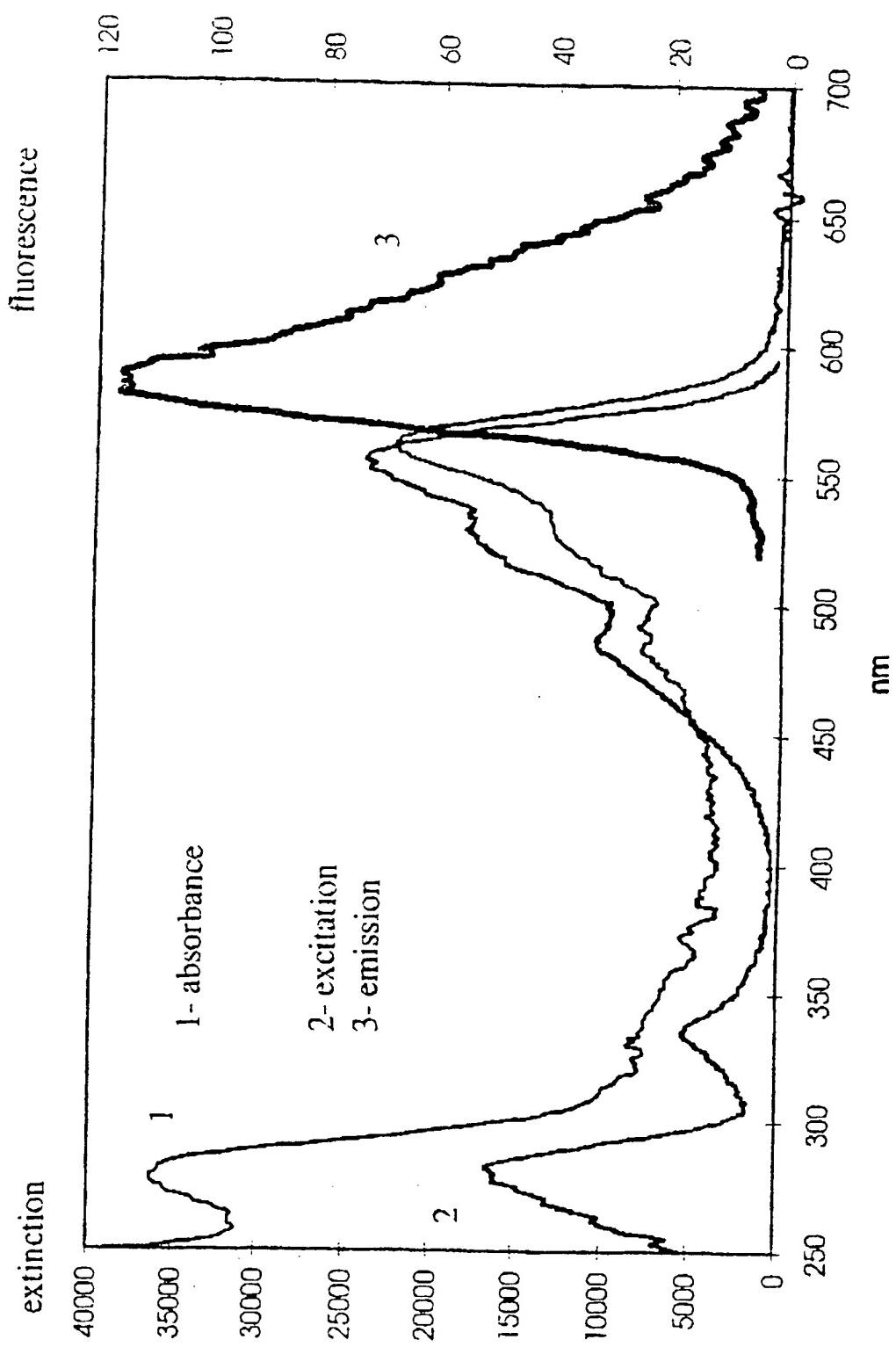


Figure 2

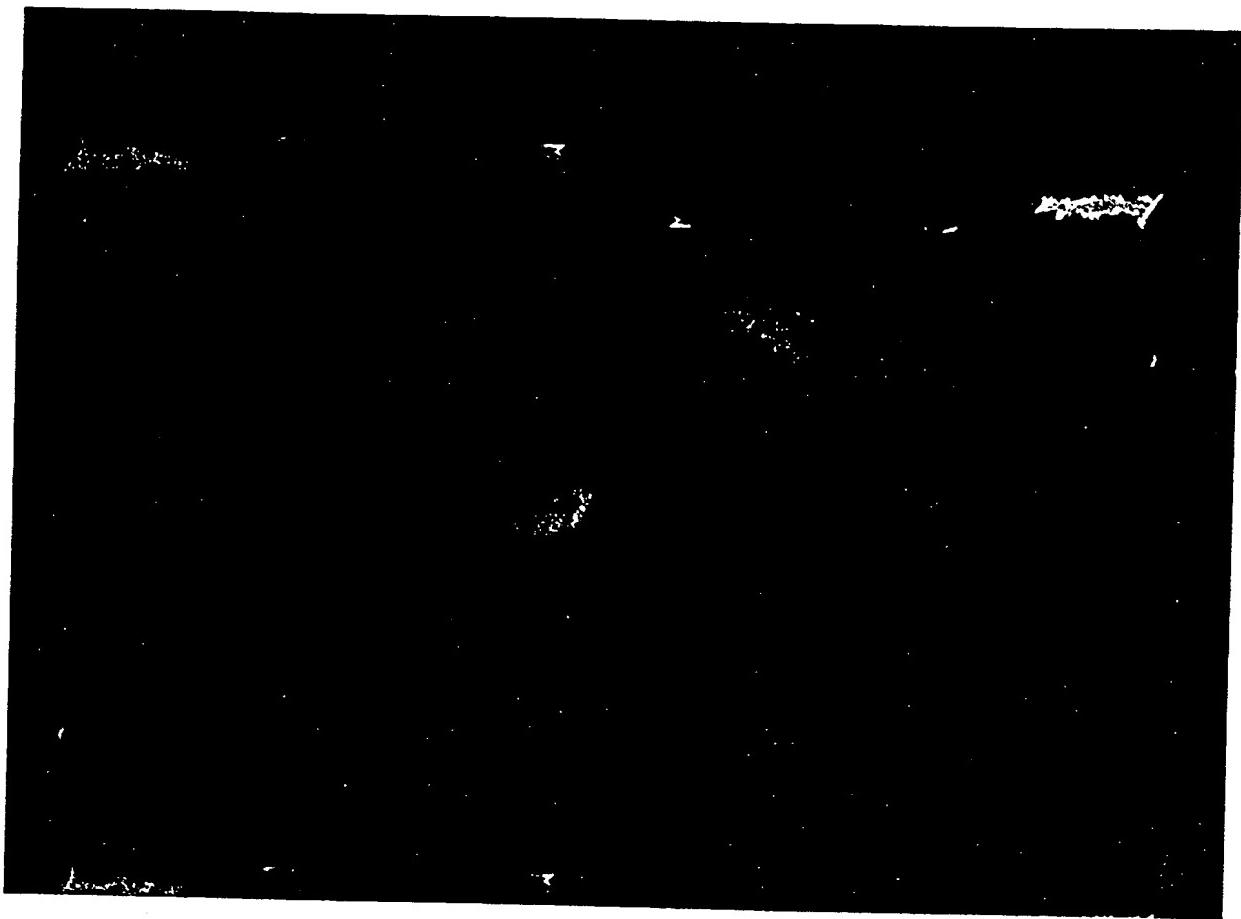


Fig. 3 A

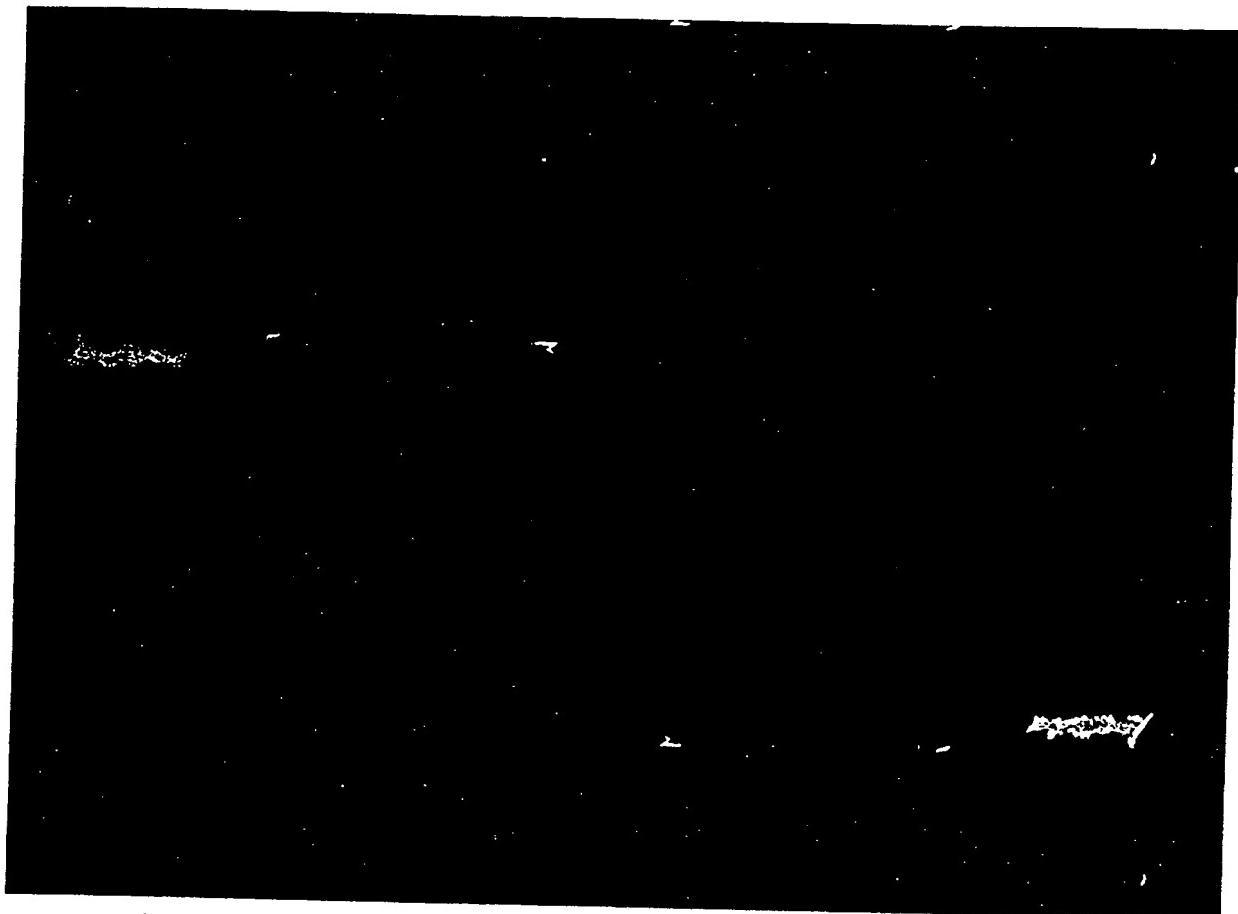


Fig. 3 B



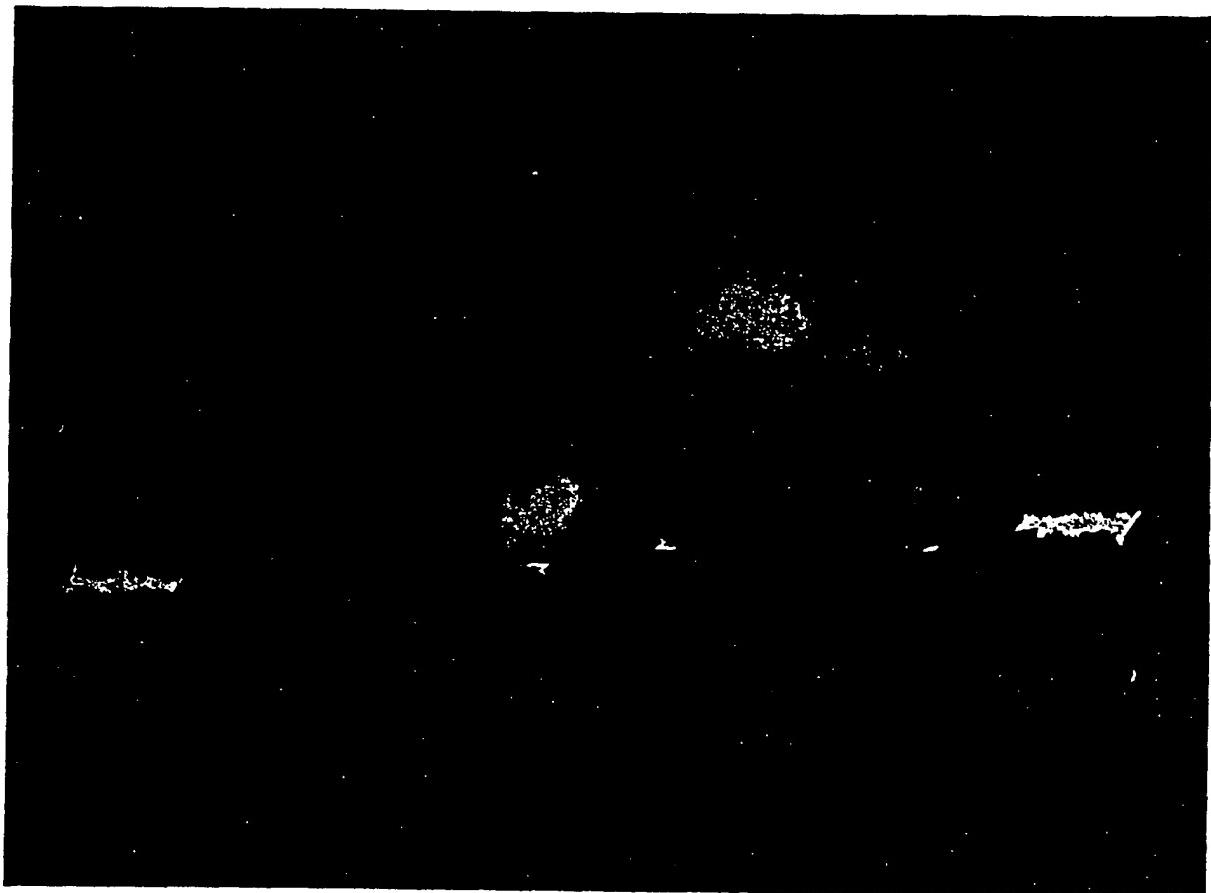


Fig. 3 C



Fig. 3 D



7/10

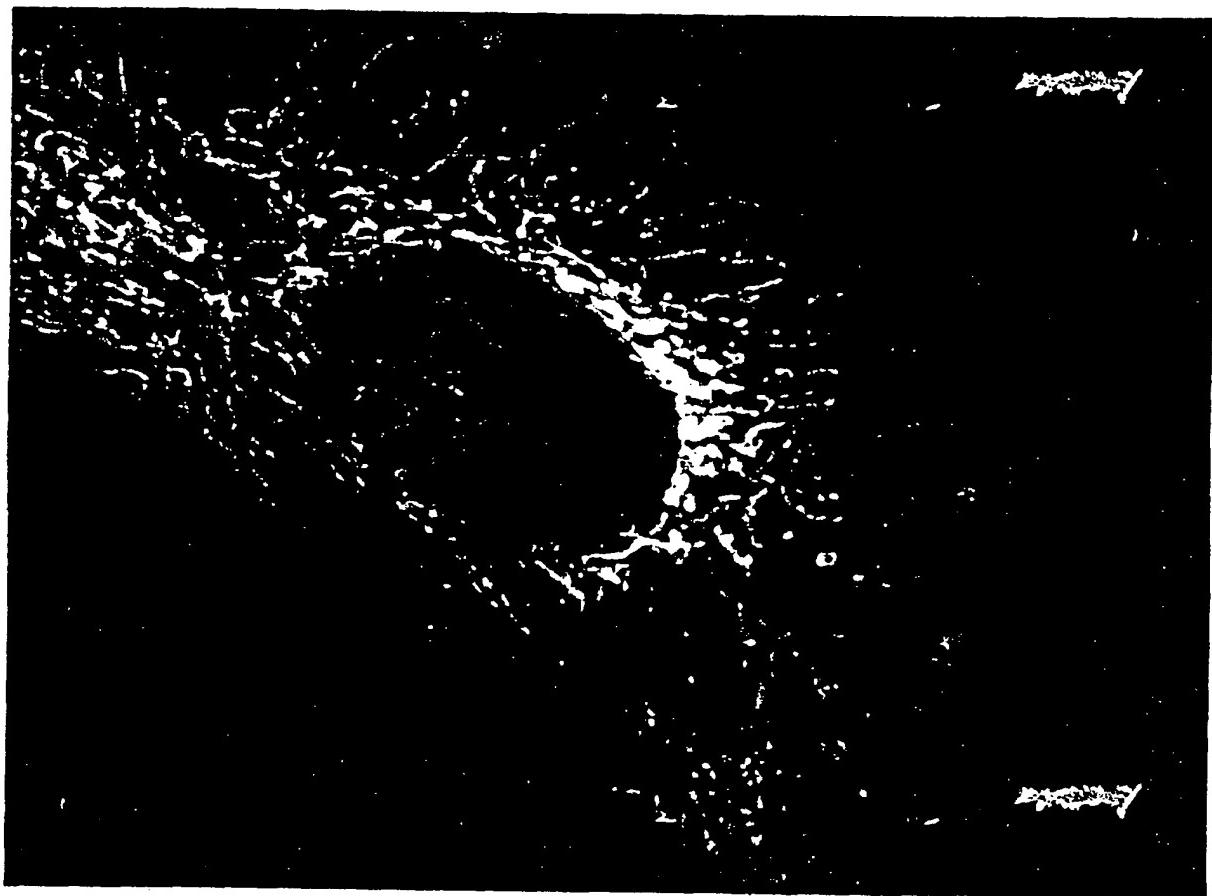


Fig. 4 F

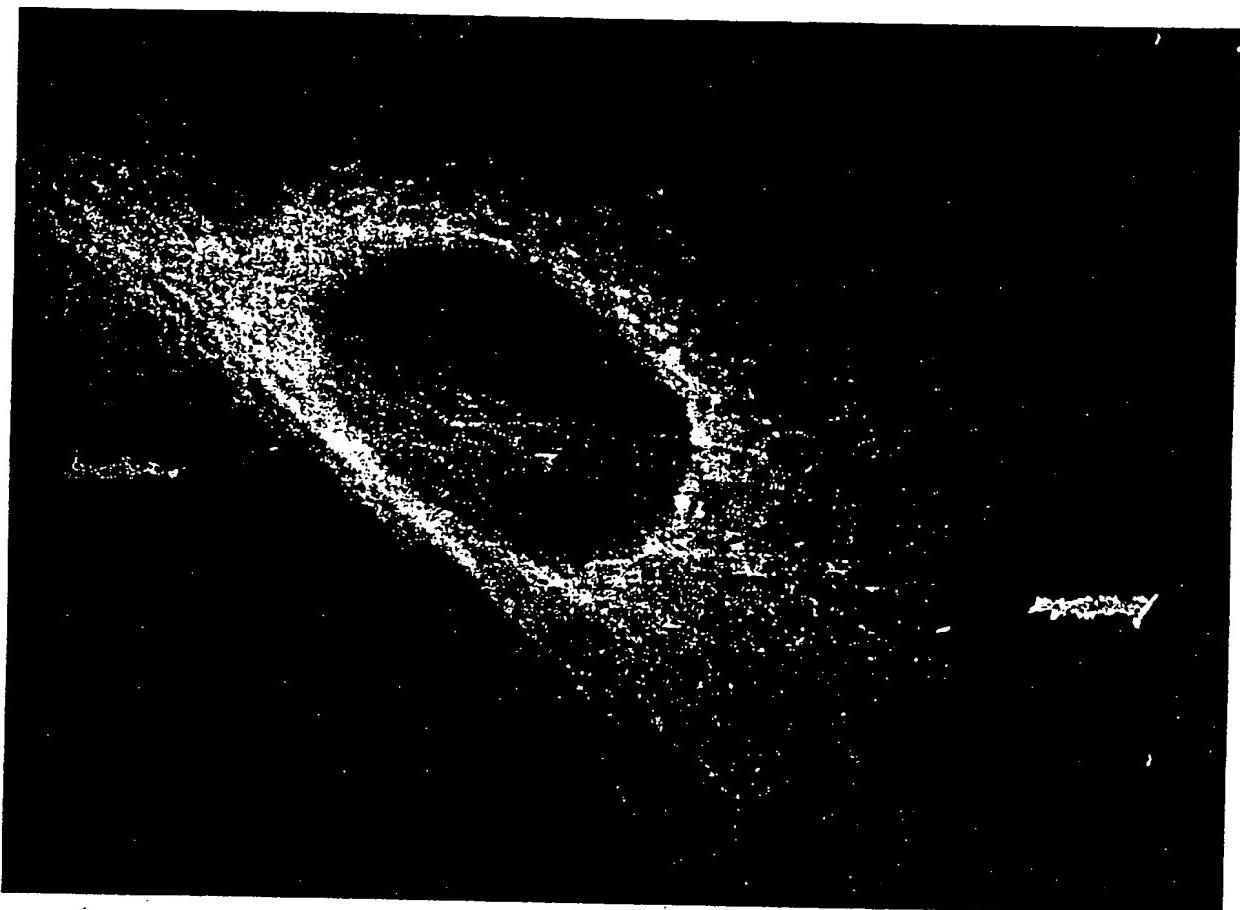


Fig. 4B

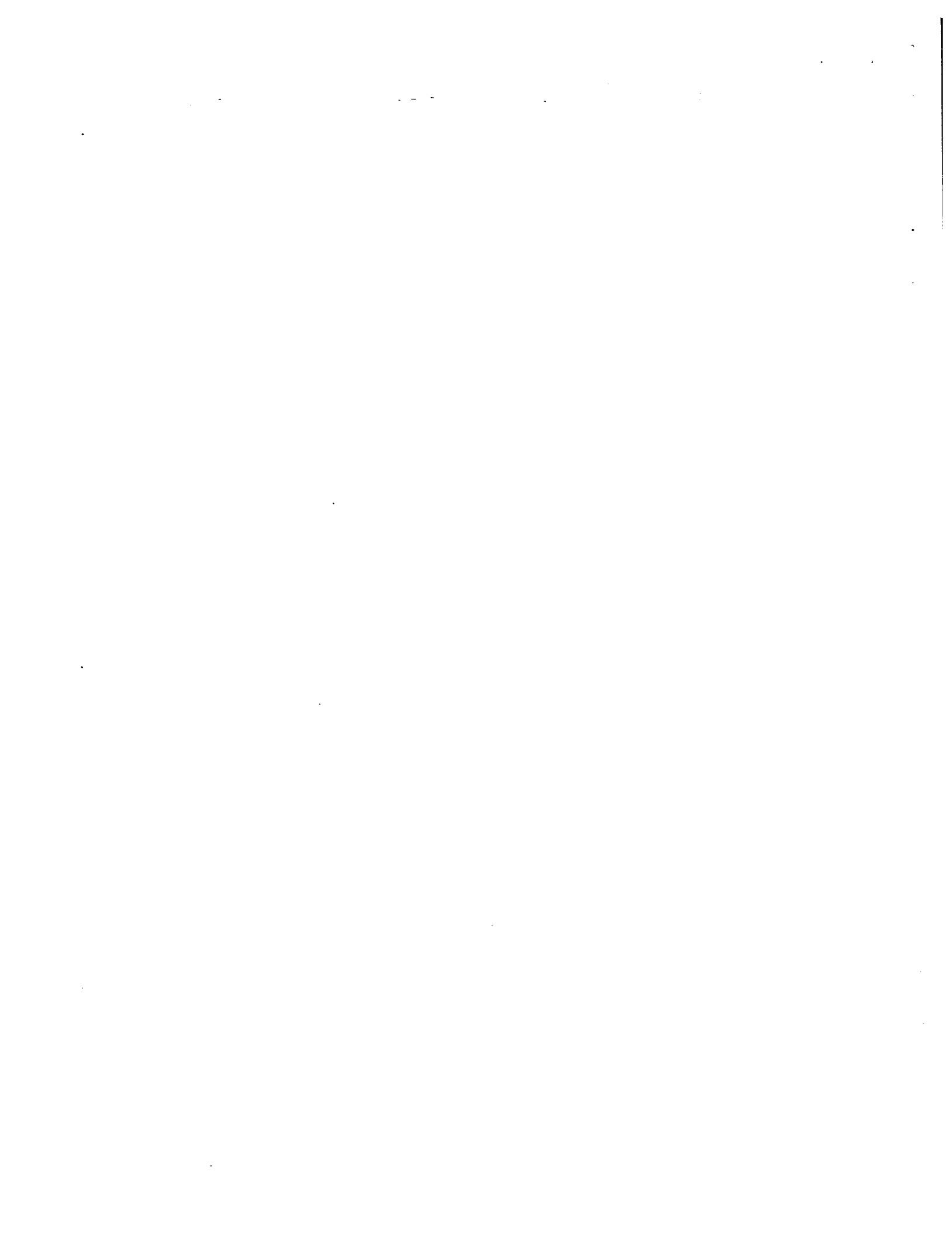




Fig. 4 c

10/10



Fig. 4 D

SEQUENCE LISTING

<110> Lukyanov, Sergey A.
Labas, Yulii A.
Matz, Mikhail V.
5 Fradkov, Arcady F.
Ding, Li
Chen, Ying
Green, Gisele
<120> Fluorescent proteins from non-bioluminescent
10 species of Class Anthozoa, genes encoding such
proteins and uses thereof
<130> D6196D6/PCT
<140> 09/418,529
<141> 1999-10-14
15 <150> 09/210,330
<151> 1998-12-11
<160> 67

<210> 1
20 <211> 25
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
25 <223> primer TN3 used in cDNA synthesis and RACE
<400> 1

cgcagtcgac cgttttttt ttttt 25

<210> 2
30 <211> 23
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
35 <223> primer TS used in cDNA synthesis and RACE
<400> 2

aagcagtggc atcaacgcag agt 23

<210> 3
<211> 6
<212> PRT
<213> *Aequorea victoria*

5 <220>
<222> 21
<223> amino acid sequence of a key stretch on which primer NGH is based; Xaa at position 21 represents unknown

10 <400> 3

Gly Xaa Val Asn Gly His
5

<210> 4
15 <211> 20
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
20 <222> 12
<223> primer NGH used for isolation of fluorescent protein; n at position 12 represents any of the four bases
<400> 4

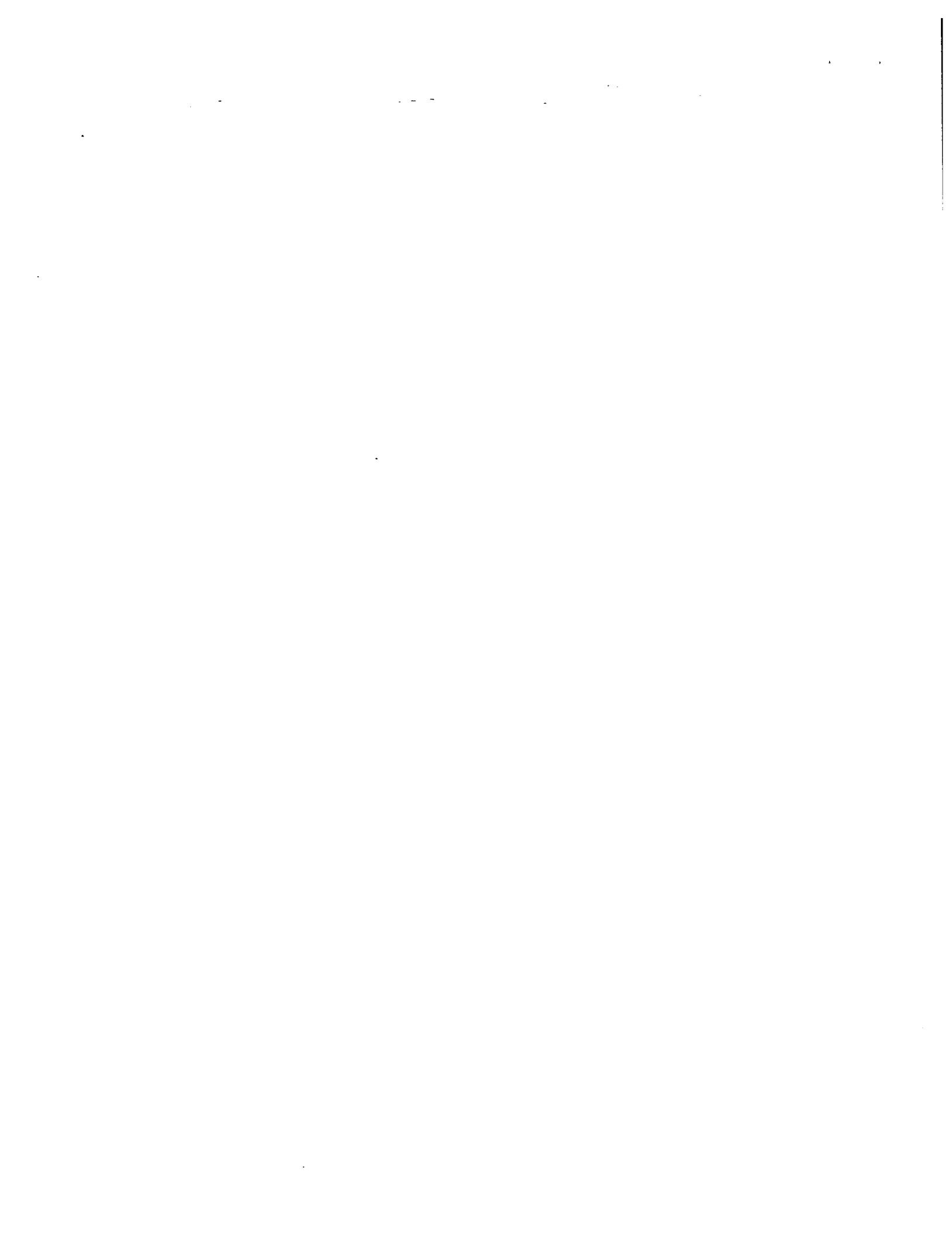
25 gayggctgcg tnaayggdca 20

<210> 5
<211> 5
<212> PRT
30 <213> *Aequorea victoria*
<220>
<222> 31...35
<223> amino acid sequence of a key stretch on which primers GEGA and GEGb are based

35 <400> 5

Gly Glu Gly Glu Gly
5

	<210>	6
	<211>	20
	<212>	DNA
5	<213>	artificial sequence
	<220>	
	<221>	primer_bind
	<223>	primer GEGa used for isolation of fluorescent protein
10	<400>	6
	gttacaggtg argggmarggg 20	
	<210>	7
	<211>	20
15	<212>	DNA
	<213>	artificial sequence
	<220>	
	<221>	primer_bind
	<223>	primer GEGb used for isolation of fluorescent protein
20	<400>	7
	gttacaggtg arggkgarggg 20	
	<210>	8
	<211>	5
25	<212>	PRT
	<213>	Aequorea victoria
	<220>	
	<222>	31..35
30	<223>	amino acid sequence of a key stretch on which primers GNGa and GNGb are based
	<400>	8
	Gly Glu Gly Asn Gly	
		5
35	<210>	9
	<211>	20



<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
5 <223> primer GNGa used for isolation of fluorescent
protein
<400> 9

gttacaggtg arggmaaygg 20

10 <210> 10
<211> 20
<212> DNA
<213> artificial sequence
<220>
15 <221> primer_bind
<223> primer GNGb used for isolation of fluorescent
protein
<400> 10

gttacaggtg arggkaaygg 20

20 <210> 11
<211> 5
<212> PRT
<213> *Aequorea victoria*

25 <220>
<222> 127..131
<223> amino acid sequence of a key stretch on which
primer NFP is based
<400> 11

30 Gly Met Asn Phe Pro
5
<210> 12
<211> 5

35 <212> PRT
<213> *Aequorea victoria*
<220>

<222> 127..131
<223> amino acid sequence of a key stretch on which
primer NFP is based
<400> 12

5 Gly Val Asn Phe Pro

5

<210> 13
<211> 20
10 <212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> primer NFP used for isolation of fluorescent
15 protein
<400> 13

ttccayggtr tgaayttycc 20

<210> 14
20 <211> 4
<212> PRT
<213> *Aequorea victoria*
<220>
<222> 134..137
25 <223> amino acid sequence of a key stretch on which
primers PVMa and PVMb are based
<400> 14

Gly Pro Val Met

30 <210> 15
<211> 21
<212> DNA
<213> artificial sequence
35 <220>
<221> primer_bind

<222> 15
 <223> primer PV_nA used for isolation of fluorescent protein; n at position 15 represents any of the four bases
 5 <400> 15
 cctgccrayg gtccngtmat g 21
 <210> 16
 <211> 21
 10 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
 <222> 15
 15 <223> primer PV_nB used for isolation of fluorescent protein; n at position 15 represents any of the four bases
 <400> 16
 cctgccrayg gtccngtmat g 21
 20 <210> 17
 <211> 47
 <212> DNA
 <213> artificial sequence
 25 <220>
 <221> primer_bind
 <223> primer T7-TN3 used in cDNA synthesis and RACE
 <400> 17
 gtaatacgac tcactatagg gccgcagtcg accgttttt tttttt
 30 <210> 18
 <211> 45
 <212> DNA
 <213> artificial sequence
 35 <220>
 <221> primer_bind
 <223> primer T7-TS used in cDNA synthesis and RACE

18

gtaatacgac tcactatagg gcaaggcagtgtatcaacgc agagt

45

19

5 22

DNA

artificial sequence

<220>

primer_bind

10 primer T7 used in cDNA synthesis and RACE

<400> 19

gtaatacgac tcactatagg gc

22

20

15 21

DNA

artificial sequence

<220>

primer_bind

20 <223> gene-specific primer used for 5'-RACE for
Anemonia majano

<400> 20

gaaatagtca ggcatactgg t

21

25 <210> 21

<211> 20

<212> DNA

<213> artificial sequence

<220>

30 <221> primer_bind

<223> gene-specific primer used for 5'-RACE for
Anemonia majano

<400> 21

gtcaggcata ctggtaggat

20

35



<210> 22
<211> 21
<212> DNA
<213> artificial sequence
5 <220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Clavularia sp.
<400> 22

10 cttgaaatag tctgctatat c 21

<210> 23
<211> 19
<212> DNA
15 <213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Clavularia sp.
20 <400> 23

tctgctatat cgtctgggt 19

<210> 24
<211> 23
25 <212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
30 <223> *Zoanthus sp.*
<400> 24

gttcttgaaa tagtctacta tgt 23

<210> 25
35 <211> 20



<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
5 <223> gene-specific primer used for 5'-RACE for
Zoanthus sp.
<400> 25

gtctactatg tcttgaggat 20

10 <210> 26
<211> 19
<212> DNA
<213> artificial sequence
<220>
15 <221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "red"
<400> 26

caagcaaatg gcaaaggtc 19

20 <210> 27
<211> 19
<212> DNA
<213> artificial sequence
25 <220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "red"
<400> 27

30 cggatttgtg gccttcgtt 19

<210> 28
<211> 19
<212> DNA
35 <213> artificial sequence
<220>

<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma striata
<400> 28

5 ttgtcttctt ctgcacaac 19

<210> 29
<211> 17
<212> DNA
10 <213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma striata
15 <400> 29

ctgcacaacg ggtccat 17

<210> 30
<211> 20
20 <212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> gene-specific primer used for 5'-RACE for
25 <213> *Anemonia sulcata*
<400> 30

cctctatctt catttcctgc 20

<210> 31
30 <211> 20
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
35 <223> gene-specific primer used for 5'-RACE for
Anemonia sulcata
<400> 31

tatcttcatt tcctgcgtac

20

5 <210> 32
 <211> 19
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
 <223> gene-specific primer used for 5'-RACE for
10 *Discosoma sp. "magenta"*
 <400> 32

ttcagcaccc catcacgag 19

15 <210> 33
 <211> 19
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
20 <223> gene-specific primer used for 5'-RACE for
 Discosoma sp. "magenta"
 <400> 33

acgctcagag ctgggttcc 19

25 <210> 34
 <211> 22
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
30 <223> gene-specific primer used for 5'-RACE for
 Discosoma sp. "green"
 <400> 34

ccctcagcaa tccatcacgt tc 22

35 <210> 35

<211> 20
<212> DNA
<213> artificial sequence
<220>
5 <221> primer_bind
<223> gene-specific primer used for 5'-RACE for
Discosoma sp. "green"
<400> 35

attatctcag tggatggttc 20
10 <210> 36
<211> 31
<212> DNA
<213> artificial sequence
15 <220>
<221> primer_bind
<223> upstream primer used to obtain full coding
region of nFPs from *Anemonia majano*
<400> 36

20 acatggatcc gctcttcaa acaagttat c 31

<210> 37
<211> 34
<212> DNA
25 <213> artificial sequence
<220>
<221> primer_bind
<223> downstream primer used to obtain full coding
region of nFPs from *Anemonia majano*
30 <400> 37

tagtactcga gcttattcgt atttcagtga aatc 34

<210> 38
<211> 29
35 <212> DNA
<213> artificial sequence
<220>

<221> primer_bind
<223> upstream primer used to obtain full coding
region of nFPs from *Clavularia sp.*
<400> 38

5 acatggatcc aacattttt tgagaaaacg 29

<210> 39
<211> 28
<212> DNA
10 <213> artificial sequence
<220>
<221> primer_bind
- <223> upstream primer used to obtain full coding
region of nFPs from *Clavularia sp.*
15 <400> 39

acatggatcc aaagctctaa ccaccatg 28

<210> 40
<211> 31
20 <212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> downstream primer used to obtain full coding
25 region of nFPs from *Clavularia sp.*
<400> 40

tagtactcga gcaacacaaa ccctcagaca a 31

<210> 41
30 <211> 28
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
35 <223> upstream primer used to obtain full coding
region of nFPs from *Zoanthus sp.*



<400> 41

acatggatcc gctcagtc aa agcacgg t 28

<210> 42

5 <211> 32

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

10 <223> downstream primer used to obtain full coding
region of nFPs from *Zoanthus sp.*

<400> 42

tagtactcg a ggttggaact acattcttat ca 32

15 <210> 43

<211> 31

<212> DNA

<213> artificial sequence

<220>

20 <221> primer_bind

<223> upstream primer used to obtain full coding
region of nFPs from *Discosoma sp.* "red"

<400> 43

acatggatcc aggtcttcca agaatgttat c 31

25

<210> 44

<211> 29

<212> DNA

<213> artificial sequence

30 <220>

<221> primer_bind

<223> downstream primer used to obtain full coding
region of nFPs from *Discosoma sp.* "red"

<400> 44

35 tagtactcg a ggagccaa gt tcagcctta 29

<210> 45
<211> 28
<212> DNA
<213> artificial sequence
5 <220>
<221> primer_bind
<223> upstream primer used to obtain full coding
region of nFPs from *Discosoma striata*
<400> 45

10 acatggatcc agttggtcca agagtgtg 28

<210> 46
<211> 28
<212> DNA
15 <213> artificial sequence
<220>
<221> primer_bind
<223> downstream primer used to obtain full coding
region of nFPs from *Discosoma striata*
20 <400> 46

tagcgagctc tatcatgcct cgtcacct 28

<210> 47
<211> 31
25 <212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> upstream primer used to obtain full coding
30 region of nFPs from *Anemonia sulcata*
<400> 47

acatggatcc gcttccttt taaagaagac t 31

<210> 48
35 <211> 28
<212> DNA
<213> artificial sequence

<220>
<221> primer_bind
<223> downstream primer used to obtain full coding
region of nFPs from *Anemonia sulcata*

5 <400> 48

tagtactcg a gtccttggga gcggcttg 28

<210> 49
<211> 30
10 <212> DNA
<213> artificial sequence
<220>
<221> primer_bind
<223> upstream primer used to obtain full coding
region of nFPs from *Discosoma sp.* "magenta"
15 <400> 49

acatggatcc agttgttcca agaatgtgat 30

<210> 50
20 <211> 26
<212> DNA
<213> artificial sequence
<220>
<221> primer_bind
25 <223> downstream primer used to obtain full coding
region of nFPs from *Discosoma sp.* "magenta"
<400> 50

tagtactcg a ggccattacg ctaatc 26

30 <210> 51
<211> 31
<212> DNA
<213> artificial sequence
<220>
35 <221> primer_bind
<223> upstream primer used to obtain full coding
region of nFPs from *Discosoma sp.* "green"



<400> 51

acatggatcc agtgcactta aagaagaaat g 31

<210> 52

5 <211> 29

<212> DNA

<213> artificial sequence

<220>

<221> primer_bind

10 <223> downstream primer used to obtain full coding
region of nFPs from *Discosoma sp.* "green"

<400> 52

tagtactcga gattcggttt aatgccttg 29

15 <210> 53

<211> 33

<212> DNA

<213> artificial sequence

<220>

20 <221> primer_bind

<223> TS-oligo used in cDNA synthesis and RACE

<400> 53

aagcagtgg atcaacgcag agtacgcrgr grg 33

25 <210> 54

<211> 238

<212> PRT

<213> *Aequorea victoria*

<220>

30 <223> amino acid sequence of GFP

<400> 54

Met Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu

5 10 15

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser

35 20 25 30

Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys

	35	40	45
	Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu		
	50	55	60
	Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro		
5	65	70	75
	Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu		
	80	85	90
	Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn		
	95	100	105
10	Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val		
	110	115	120
	Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn		
	—	125	130
	Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val		
15	140	145	150
	Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe		
	155	160	165
	Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp		
	170	175	180
20	His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu		
	185	190	195
	Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp		
	200	205	210
	Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr		
25	215	220	225
	Ala Ala Gly Ile Thr His Gly Met Asp Glu Leu Tyr Lys		
	230	235	
	<210>	55	
30	<211>	898	
	<212>	DNA	
	<213>	<i>Discosoma</i> sp. "red"	
	<220>		
	<221>	CDS	
35	<223>	cDNA sequence of drFP583	
	<400>	55	
	gtcctcccaa gcagtggtat caacgcagag tacggggag tttcagccag	50	
	tgacggtcag tgacagggtg agccacttgg tataccaaca aaatgaggtc	100	



ttccaagaat gttatcaagg agttcatgag gtttaagggtt cgtatggaag 150
 gaacggtaaa tgcccacgag tttgaaatag aaggcgaagg agaggggagg 200
 ccatacgaag gccacaatac cgtaaagctt aaggtAACCA aggggggacc 250
 tttgccattt gcttggata ttttgcacc acaatttcag tatggaagca 300
 5 aggtatatgt caagcacccct gccgacatac cagactataa aaagctgtca 350
 tttcctgaag gatttaaatg ggaaagggtc atgaactttg aagacggtgg 400
 cgtcgttact gtaacccagg attccagttt gcaggatggc tgtttcatct 450
 acaagtcaag ttcattggcg ttgaacttcc cttccgatgg acctgttatg 500
 caaaaagaaga caatgggctg ggaagccagc actgagcggt tttatcctcg 550
 10 tggatggcgtg ttgaaaggag agattcataa ggctctgaag ctgaaagacg 600
 gtggtcatta cctagttgaa ttcaaaagta tttacatggc aaagaagcct 650
 gtgcagctac cagggtacta ctatgttgc tccaaactgg atataacaag 700
 ccacaacgaa gactatacaa tcgttgagca gtatgaaaga accgagggac 750
 gccaccatct gttcccttaa ggctgaactt ggctcagacg tgggtgagcg 800
 15 gtaatgacca caaaaggcag cgaagaaaaa ccatgatcgt ttttttagg 850
 ttggcagcct gaaatcgtag gaaatacatc agaaatgtta caaacagg 898

	<210>	56	
	<211>	225	
20	<212>	PRT	
	<213>	<i>Discosoma sp. "red"</i>	
	<220>		
	<223>	amino acid sequence of drFP583	
	<400>	56	
25	Met Arg Ser Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys		
	5	10	15
	Val Arg Met Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Glu		
	20	25	30
	Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly His Asn Thr Val Lys		
30	35	40	45
	Leu Lys Val Thr Lys Gly Pro Leu Pro Phe Ala Trp Asp Ile		
	50	55	60
	Leu Ser Pro Gln Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His		
	65	70	75
35	Pro Ala Asp Ile Pro Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly		
	80	85	90
	Phe Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Val Val		

	95	100	105
	Thr Val Thr Gln Asp Ser Ser Leu Gln Asp Gly Cys Phe Ile Tyr		
	110	115	120
	Lys Val Lys Phe Ile Gly Val Asn Phe Pro Ser Asp Gly Pro Val		
5	125	130	135
	Met Gln Lys Lys Thr Met Gly Trp Glu Ala Ser Thr Glu Arg Leu		
	140	145	150
	Tyr Pro Arg Asp Gly Val Leu Lys Gly Glu Ile His Lys Ala Leu		
	155	160	165
10	Lys Leu Lys Asp Gly Gly His Tyr Leu Val Glu Phe Lys Ser Ile		
	170	175	180
	Tyr Met Ala Lys Lys Pro Val Gln Leu Pro Gly Tyr Tyr Tyr Val		
	185	190	195
	Asp Ser Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile		
15	200	205	210
	Val Glu Gln Tyr Glu Arg Thr Glu Gly Arg His His Leu Phe Leu		
	215	220	225

	<code><210></code>	57
20	<code><211></code>	695
	<code><212></code>	DNA
	<code><213></code>	artificial sequence
	<code><220></code>	
	<code><223></code>	nucleotide sequence of humanized drFP583
25	<code><400></code>	57

	ggggtaacca tgcgctcctc caagaacgtc atcaaggagt tcatgcgctt 50
	caaggtgcgc atggaggggca ccgtgaacgg ccacgagttc gagatcgagg 100
	gcgagggcga gggccgcccc tacgagggcc acaacaccgt gaagctgaag 150
	gtgaccaagg gcggcccccct gcccttcgcc tgggacatcc tgtccccca 200
30	gttccagtac ggctccaagg tgtacgtcaa gcaccccgcc gacatccccg 250
	actacaagaa gctgtccttc cccgagggtc tcaagtggga ggcgtgtatg 300
	aacttcgagg acggcggcgt ggtgaccgtg acccaggact cctccctgca 350
	ggacggctgc ttcatctaca aggtgaagtt catcggcgtg aacttcccct 400
	ccgacggccc cgtatgcag aagaagacca tgggctggga ggctccacc 450
35	gagcgcctgt acccccgcga cggcgtgtc aagggcgaga tccacaaggc 500
	cctgaagctg aaggacggcg gccactacct ggtggagttc aagtccatct 550
	acatggccaa gaagcccggtg cagctgccccg gctactacta cgtggactcc 600

aagctggaca tcacctccca caacgaggac tacaccatcg tggaggcagta 650
cgagcgcacc gagggccgcc accacctgtt cctgtaatct agagc 695

5 8<210> 58
 <211> 31
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
10 <223> oligonucleotide primer used for amplifying
 non-humanized wild type coding region
 fragments from drFP583 and dmFP592
 — <400> 58

15 acatggatcc aggtcttcca agaatgttat c 31
 <210> 59
 <211> 26
 <212> DNA
 <213> artificial sequence
20 <220>
 <221> primer_bind
 <223> oligonucleotide primer used for amplifying
 non-humanized wild type coding region
 fragments from drFP583 and dmFP592
25 <400> 59

tagtactcga gccaaagtca gcctta 26

30 <210> 60
 <211> 30
 <212> DNA
 <213> artificial sequence
 <220>
 <221> primer_bind
 <223> oligonucleotide primer used for amplifying
 non-humanized wild type coding region
 fragments from drFP583 and dmFP592
35 <400> 60

	acatggatcc agttgttcca agaatgtgat	30
	<210> 61	
	<211> 27	
5	<212> DNA	
	<213> artificial sequence	
	<220>	
	<221> primer_bind	
10	<223> oligonucleotide primer used for amplifying non-humanized wild type coding region fragments from drFP583 and dmFP592	
	<400> 61	
	tagt a ctcgaa gcccattacc gctaatac	27
15	<210> 62	
	<211> 678	
	<212> DNA	
	<213> artificial sequence	
	<220>	
20	<223> nucleotide sequence of humanized drFP583/dmFP592	
	<400> 62	
	atgagctgca gcaagaacgt gatcaaggag ttcatgcggc tcaagggtgcg 50	
	gatggagggc accgtgaacg gccacgagtt cgagatcaag ggcgagggcg 100	
25	agggccggcc ctacgagggc cactgcagcg tgaagctcat ggtgaccaag 150	
	ggcggcccccc tccccttcgc cttcgacatc ctcagcccccc agttccagta 200	
	cggcagcaag gtgtacgtga agcaccccgcc cgacatcccc gactacaaga 250	
	agctcagctt ccccgagggc ttcaagtggg agcgggtgat gaacttcgag 300	
	gacggcggcg tggtgaccgt gagccaggac agcagcctca aggacggctg 350	
30	cttcatctac gaggtgaagt tcatcggcgt gaacttcccc agcgacggcc 400	
	ccgtgatgca gcggcggacc cggggctggg aggccagcag cgagcggctc 450	
	taccccccggg acggcgtgct caagggcgac atccacatgg ccctccggct 500	
	cgagggcggc ggccactacc tcgtggagtt caagagcatc tacatggcca 550	
	agaagccccgt gcagctcccc ggctactact acgtggacag caagctcgac 600	
35	atcaccagcc acaacgagga ctacaccatc gtggagcagt acgagcggac 650	
	cgagggccgg caccacctct tcctctga 678	
	<210> 63	



<211> 225
<212> PRT
<213> artificial sequence
<220>
5 <223> amino acid sequence of humanized
drFP583/dmFP592
<400> 63

	Met Ser Cys Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Lys		
	5	10	15
10	Val Arg Met Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Lys		
	20	25	30
	Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly His Cys Ser Val Lys		
	— 35	40	45
	Leu Met Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile		
15	50	55	60
	Leu Ser Pro Gln Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His		
	65	70	75
	Pro Ala Asp Ile Pro Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly		
	80	85	90
20	Phe Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val		
	95	100	105
	Thr Val Ser Gln Asp Ser Ser Leu Lys Asp Gly Cys Phe Ile Tyr		
	110	115	120
	Glu Val Lys Phe Ile Gly Val Asn Phe Pro Ser Asp Gly Pro Val		
25	125	130	135
	Met Gln Arg Arg Thr Arg Gly Trp Glu Ala Ser Ser Glu Arg Leu		
	140	145	150
	Tyr Pro Arg Asp Gly Val Leu Lys Gly Asp Ile His Met Ala Leu		
	155	160	165
30	Arg Leu Glu Gly Gly His Tyr Leu Val Glu Phe Lys Ser Ile		
	170	175	180
	Tyr Met Ala Lys Lys Pro Val Gln Leu Pro Gly Tyr Tyr Val		
	185	190	195
	Asp Ser Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile		
35	200	205	210
	Val Glu Gln Tyr Glu Arg Thr Glu Gly Arg His His Leu Phe Leu		
	215	220	225

<210> 64
<211> 678
<212> DNA
<213> artificial sequence
5 <220>
<223> nucleotide sequence of drFP583/dmFP592-2G
<400> 64
atgagctgca gcaagaacgt gatcaaggag ttcatgcggc tccaggtgcg 50
gatggagggc accgtgaacg gccacgagtt cgagatcaag ggcgagggcg 100
10 agggccggcc ctacgagggc cactgcagcg tgaagctcat ggtgaccaag 150
ggcggcccccc tccccttcgc cttcgacatc ctcagcccccc agttccagta 200
cggcagcaag gtgtacgtga agcacccgc cgacatcccc gactacaaga 250
agctcagctt ccccgagggc ttcaagtggg agcgggtgat gaacttcgag 300
gacggcggcg tggtgaccgt gagccaggac agcagcctca aggacggctg 350
15 cttcatctac gaggtgaagt tcatggcgt gaacttcccc agcgacggcc 400
ccgtgatgca gcggcggacc cggggctggg aggccagcag cgagcggctc 450
taccccccggg acggcgtgct caagggcgac atccacatgg ccctccggct 500
cgagggcggc ggccactacc tcgtggagtt caagagcatc tacatggcca 550
agaagccgt gcagctcccc ggctactact acgtggacag caagctcgac 600
20 atcaccagcc acaacgagga ctacaccatc gtggagcagt acgagcggtc 650
cgagggccgg caccacctct tcctctga 678

<210> 65
25 <211> 225
<212> PRT
<213> artificial sequence
<220>
<223> amino acid sequence of drFP583/dmFP592-2G
30 <400> 65
Met Ser Cys Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Gln
5 10 15
Val Arg Met Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Lys
20 25 30
35 Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly His Cys Ser Val Lys
35 40 45
Leu Met Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile
50 55 60

	Leu Ser Pro Gln Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His		
	65	70	75
	Pro Ala Asp Ile Pro Asp Tyr Lys Lys Leu Ser Phe Pro Glu Gly		
	80	85	90
5	Phe Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val		
	95	100	105
	Thr Val Ser Gln Asp Ser Ser Leu Lys Asp Gly Cys Phe Ile Tyr		
	110	115	120
	Glu Val Lys Phe Ile Gly Val Asn Phe Pro Ser Asp Gly Pro Val		
10	125	130	135
	Met Gln Arg Arg Thr Arg Gly Trp Glu Ala Ser Ser Glu Arg Leu		
	140	145	150
	Tyr Pro Arg Asp Gly Val-Leu Lys Gly Asp Ile His Met Ala Leu		
	155	160	165
15	Arg Leu Glu Gly Gly His Tyr Leu Val Glu Phe Lys Ser Ile		
	170	175	180
	Tyr Met Ala Lys Lys Pro Val Gln Leu Pro Gly Tyr Tyr Tyr Val		
	185	190	195
	Asp Ser Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile		
20	200	205	210
	Val Glu Gln Tyr Glu Arg Ser Glu Gly Arg His His Leu Phe Leu		
	215	220	225

	<210>	66
25	<211>	678
	<212>	DNA
	<213>	artificial sequence
	<220>	
	<223>	nucleotide sequence of drFP583/dmFP592-Q3
30	<400>	66
	atgagctgca gcaagaacgt gatcaaggag ttcatgcggc tccaggtgcg	50
	gatggagggc accgtgaacg gccacgagtt cgagatcaag ggcgagggcg	100
	agggccggcc ctacgagggc cactgcagcg tgaagctcat ggtgaccaag	150
	ggcggcccccc tccccttcgc cttcgacatc ctcagcccc agtccagta	200
35	cggcagcaag gtgtacgtga agcacccgc cgacatcccc gactacatga	250
	agctcagctt ccccgagggc ttcaagtggg agcgggtgat gaacttcgag	300
	gacggcggcg tggtgaccgt gagccaggac agcagcctca aggacggctg	350



cttcatctac gaggtgaagt tcatacgccgt gaactcccc agcgacggcc 400
ccgtgatgca gcggcggacc cggggctggg aggccagcag cgagcggctc 450
taccccccggg acggcgtgct caagggcgac atccacatgg ccctccggct 500
cgagggcggc ggccactacc tcgtggagtt caagagcatc tacatggcca 550
agaagcccggt gcagctcccc ggctactact acgtggacag caagctcgac 600
atcaccagcc acaacgagga ctacaccatc gtggagcagt acgagcggtc 650
cgagggccgg caccacctct tcctctga 678

	<210>	67	
10	<211>	225	
	<212>	PRT	
	<213>	artificial sequence	
	<220>	-	
	<223>	amino acid sequence of drFP583/dmFP592-Q3	
15	<400>	67	
	Met Ser Cys Ser Lys Asn Val Ile Lys Glu Phe Met Arg Phe Gln		
	5	10	15
	Val Arg Met Glu Gly Thr Val Asn Gly His Glu Phe Glu Ile Lys		
	20	25	30
20	Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly His Cys Ser Val Lys		
	35	40	45
	Leu Met Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Phe Asp Ile		
	50	55	60
	Leu Ser Pro Gln Phe Gln Tyr Gly Ser Lys Val Tyr Val Lys His		
25	65	70	75
	Pro Ala Asp Ile Pro Asp Tyr Met Lys Leu Ser Phe Pro Glu Gly		
	80	85	90
	Phe Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val		
	95	100	105
30	Thr Val Ser Gln Asp Ser Ser Leu Lys Asp Gly Cys Phe Ile Tyr		
	110	115	120
	Glu Val Lys Phe Ile Gly Val Asn Phe Pro Ser Asp Gly Pro Val		
	125	130	135
	Met Gln Arg Arg Thr Arg Gly Trp Glu Ala Ser Ser Glu Arg Leu		
35	140	145	150
	Tyr Pro Arg Asp Gly Val Leu Lys Gly Asp Ile His Met Ala Leu		
	155	160	165



Arg Leu Glu Gly Gly His Tyr Leu Val Glu Phe Lys Ser Ile
170 175 180
Tyr Met Ala Lys Lys Pro Val Gln Leu Pro Gly Tyr Tyr Tyr Val
185 190 195
5 Asp Ser Lys Leu Asp Ile Thr Ser His Asn Glu Asp Tyr Thr Ile
200 205 210
Val Glu Gln Tyr Glu Arg Ser Glu Gly Arg His His Leu Phe Leu
215 220 225

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/29473

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : [IPC 7]: C07K 14/435; C12N 1/00, 1/10, 5/10, 15/12, 15/63
US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1, 252.3, 252.33, 324, 410, 254.11, 348, 369, 69.1; 530/350; 536/23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	MATZ et al. Fluorescent proteins from nonbioluminescent Anthozoa species. Nature Biotechnology. October 1999, Volume 17, No. 10, pages 969-973, see entire document.	1-24
X, P	DE 197 18 640 A1 (WIEDENMANN) 22 July 1999 (22.07.99), see entire document.	15, 20
A	US 5,491,084 A (CHALFIE et al) 13 February 1996 (13.02.96).	15, 20
X	ANDERLUH et al. Cloning, sequencing , and expression of equinatoxin II. Biochemical and Biophysical Research Communications. 1996, Volume 220, No. 2, pages 437-442, see entire document.	1,6, 9, 11-15, 20

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	Document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

09 MARCH 2000

Date of mailing of the international search report

18 APR 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Authorized officer

GABRIELE ELISABETH BUGAISKY

Faxsimile No. (703) 305-3230

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/29473

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X L	MACEK et al. Intrinsic tryptophan fluorescence of equinatoxin II, a pore-forming polypeptide from the sea anemone, <i>Actinia equina</i> L, monitors its interaction with lipid membranes. European Journal of Biochemistry. 1995, Volume 234, pages 329-335, entire document. Cited as "L" document because it establishes fluorescence of equinatoxin II.	15, 20 1, 6, 9, 11-14

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/US99/29473**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: 6-8, 10, 20-24 because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

Since the sequence diskette (CRF) submitted by applicant is defective, a sequence search could not be performed. Accordingly, claims 6-8, 10 and 20-24 were searched only in-part, based on a word search.

3. Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
 No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/29473

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

435/320.1, 252.3, 252.33, 324, 410, 254.11, 348, 369, 69.1; 530/350; 536/23.5

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

Dialog files 155, 5, 434, 34, 357, 28, 44, 35, 77(Medline, Biosis, Scisearch, Derwent Biotech Abs., Oceanic Abs., Aquatic Sci. & Fish Abs., Dissertation Abs. Online, Conference Papers Index); STN-CAS files registry, CAPLUS; WEST files USPT, Derwent WPI
search terms: fluoresc? bioluminesc? protein? polypeptide?, anthoz?, zoanthar?, corallimorph?, discosom?, rhodact?
coral? cnidar?, anemon? alga, algae, invert? coelenter?, drfp583, mrrssknvik/sqsp, vngh/sqep, gegeg/sqep, gegng/sqep,
gmnfp/sqep, gvnfp/sqep, gpvm/sqep